研 究 生 教 材
YANJIUSHENG JIAOCAI

矩阵理论及其应用

邱启荣　主编



VISIT…

LANZAROTE Caliente.com

研 究 生 教 材
YANJIUSHENG JIAOCAI

矩阵理论及其应用

本教材根据国家教委制定的工科研究生学习矩阵论课程的基本要求，较全面、系统地介绍了与工程技术联系密切、应用广泛的矩阵理论与方法。编写过程中力求做到深入浅出、简明易懂、深度与广度适中。教材中给出了丰富的例子，使读者能更好地理解抽象的概念、理论，应用实例丰富了教材内容，也可扩大学生的知识面。教材注意用新观点、新思想来审视和阐述经典内容。教材中用Matlab软件处理计算问题，一方面可以加深对问题的理解，另一方面也可使读者从繁琐的重复运算中解脱出来，也便于读者将所学知识用于工程计算。

2008

ISBN 978-7-5083-7626-4
9 787508 376264 >
定价：19.60元

0151.21/47
2008

研 究 生 教 材
YANJIUSHENG JIAOCAI

矩阵理论及其应用

主　编　邱启荣
编　写　徐英凯　孙淑珍　李国东
主　审　朱来义

中国电力出版社
http://jc.cepp.com.cn

内 容 提 要

本书对矩阵理论的基础知识作了详细的介绍，同时结合使用 Matlab 软件解决矩阵论有关内容的计算问题。全书分为八章，内容包括线性空间、线性变换、Jordan 标准形、向量与矩阵的范数、矩阵分析、矩阵函数及其应用、矩阵的分解、广义逆矩阵。本书内容资料丰富、论述详尽严谨、文字通俗易懂、便于自学。

本书可作为理工科院校硕士研究生矩阵理论课程的教材，还可作为学习矩阵理论人员的参考用书。

图书在版编目（CIP）数据

矩阵理论及其应用/邱启荣主编．—北京：中国电力出版社，2008

研究生教材

ISBN 978-7-5083-7626-4

Ⅰ．矩…　Ⅱ．邱…　Ⅲ．矩阵—理论—研究生—教材
Ⅳ．O151.21

中国版本图书馆 CIP 数据核字（2008）第 091857 号

中国电力出版社出版、发行
（北京三里河路 6 号　100044　http://jc.cepp.com.cn）
北京市同江印刷厂印刷
各地新华书店经售

*

2008 年 7 月第一版　2008 年 7 月北京第一次印刷
787 毫米×1092 毫米　16 开本　12.25 印张　297 千字
定价 **19.60** 元

敬 告 读 者

本书封面贴有防伪标签，加热后中心图案消失
本书如有印装质量问题，我社发行部负责退换

版 权 专 有　翻 印 必 究

前　言

矩阵理论既是学习经典数学的基础，又是一门最有实用价值的数学理论。它不仅是数学的一个重要的分支，而且业已成为现代各科技领域处理大量有限维空间形式与数量关系的强有力的工具。特别是计算机的广泛应用，为矩阵论的应用开辟了广阔的前景。例如，数值计算、优化方法及稳定性理论等，都与矩阵论有着密切的联系，从而使矩阵理论近年来在内容上有相当大的更新，非现有教科书所能概括的。《矩阵理论及其应用》是工科院校硕士研究生重要的学位课程，我们根据各学科研究生学习和科研的需要，结合多年的教学经验，以读者熟悉的线性代数为基础，结合 Matlab 软件应用，编出此书。

本书是编者在原教材《矩阵理论及其应用》（中国电力出版社出版，2003 年）基础上改编的，书中的大部分内容在原教材中都有。本书在兼顾矩阵理论讨论的同时，注重矩阵理论应用的讨论，加强了与线性代数有关内容的衔接，使读者更容易理解和掌握，同时为弥补现有矩阵论教材应用实例较少的不足，重新配置例题与习题，增加了使用 Matlab 软件解决在矩阵论有关内容中的计算问题的内容，适当增加了一些有关的应用实例。

全书分为八章。其中：第一、二章由徐英凯编写；第三、六章由邱启荣编写；第四、五章由孙淑珍编写；第七、八章由李国东编写。全书由邱启荣统稿并担任主编。由朱来义担任主审。

本书的编写力求做到资料丰富、论述详尽严谨、文字通俗易懂，便于自学；尽可能满足不同专业工科及理科研究生学习的需要。全书内容充实，不同专业的研究生可根据需要删减。

由于编者的水平所限，错漏和不妥之处在所难免，敬请批评指正。

编　者

2008 年 4 月 30 日

目 录

前言
第一章 线性空间…… 1
§1.1 集合与映射 …… 1
§1.2 线性空间及其性质…… 4
§1.3 基、维数与坐标…… 6
§1.4 线性子空间 …… 15
§1.5 内积空间…… 22
习题1 …… 31
第二章 线性变换 …… 34
§2.1 线性变换的定义 …… 34
§2.2 线性变换的运算 …… 38
§2.3 线性变换与矩阵 …… 39
§2.4 正交变换与正交矩阵 …… 44
§2.5 对称变换与对称矩阵 …… 46
§2.6 特征值与特征向量 …… 49
习题2 …… 57
第三章 Jordan标准形 …… 60
§3.1 λ-矩阵…… 60
§3.2 不变因子与初等因子 …… 64
§3.3 Jordan 标准形 …… 71
§3.4 Cayley-Hamilton 定理 最小多项式 …… 84
习题3 …… 91
第四章 向量与矩阵的范数 …… 95
§4.1 向量范数…… 95
§4.2 矩阵的范数 …… 98
习题4 …… 105
第五章 矩阵分析…… 107
§5.1 矩阵序列的极限…… 107
§5.2 矩阵级数 …… 109
§5.3 矩阵的Kronecker积 …… 111
§5.4 矩阵的微分和积分…… 115
习题5 …… 117
第六章 矩阵函数及其应用…… 119
§6.1 矩阵幂级数 …… 119

§6.2 矩阵函数 …… 120
§6.3 矩阵函数的一般定义及其计算 …… 125
§6.4 矩阵函数的应用 …… 129
习题 6 …… 137
第七章 矩阵的分解 …… 139
§7.1 n 阶矩阵的 LU 分解 …… 139
§7.2 矩阵的 QR 分解 …… 149
§7.3 矩阵的满秩分解 …… 157
§7.4 矩阵的奇异值分解 …… 159
习题 7 …… 162
第八章 广义逆矩阵 …… 164
§8.1 广义逆矩阵及其分类 …… 164
§8.2 广义逆矩阵 A^- …… 165
§8.3 自反广义逆 …… 171
§8.4 广义逆 A_m^- …… 173
§8.5 广义逆 A_l^- …… 176
§8.6 广义逆 A^+ …… 180
习题 8 …… 189
参考文献 …… 190

第一章 线 性 空 间

线性空间既是代数学的基本概念，也是矩阵理论的基本概念之一，本章首先介绍这一概念. 学习过这一部分内容的读者可以将本章略过，或仅作为回顾浏览.

§1.1 集 合 与 映 射

人们经常使用“一组”、“一队”、“一批”这样的词汇来描述某类事物，它们被用来表示一定事物的集体. 在数学上称它们为**集合或集**. 组成集合的内容称为集合中的**元素**.

我们用大写字母 A，B，C，… 表示集合，用小写字母 a，b，c，…表示集合的元素.

如果 a 是集合 A 的元素，就称 a 属于 A，记作 $a\in A$；如果 a 不是集合 A 的元素，就称 a 不属于 A，记作 $a\notin A$.

若集合 A 中仅含有有限个元素，则称其为有限集合，比如由 -1，1 这两个元素组成的集合就是一个有限集合.

若集合 A 中含有无限个元素，则称其为无限集合，比如由全体实数组成的集合就是一个无限集合.

集合可以用列出其所含有的全部元素，或者给出集合中元素所具备的特征的方式来表示. 当然前一种表示方式仅对有限集合适用，而无限集合的表示要用第二种方式.

例如，由 -1，1 这两个元素组成的集合可以表示为 $A=\{-1, 1\}$，而由全体实数组成的集合可以表示为 $A=\{x\,|\,x\in\mathbf{R}\}$，当然第一个集合也可以表示为 $A=\{x\,|\,x^2-1=0, x\in\mathbf{R}\}$.

不含有任何元素的集合称为**空集合**，记为 Φ. 例如集合 $A=\{x\,|\,x^2+1=0, x\in\mathbf{R}\}$就是一个空集合.

为了方便起见，我们约定：

$\mathbf{N}$ 为全体自然数组成的集合；

$\mathbf{Z}$ 为全体整数组成的集合；

$\mathbf{Q}$ 为全体有理数组成的集合；

$\mathbf{R}$ 为全体实数组成的集合；

$\mathbf{C}$ 为全体复数组成的集合.

设 A，B 是两个集合，如果 A 的每一个元素都是 B 的元素，则称 A 是 B 的子集，记作 $A\subseteq B$，或记作 $B\supseteq A$.

根据这样的定义，一个集合 A 是它自身的子集，同时规定空集合 Φ 是任意集合的子集，这两个子集称为 A 的当然子集，而 A 的其他子集称为 A 的真子集.

设 A，B 是两个集合，如果 $A\subseteq B$，并且 $B\subseteq A$，则称这两个集合相等，记作 $A=B$.

设 A，B 是两个集合，由 A 的所有元素和 B 的所有元素组成的集合称为 A 与 B 的**并集**，记作 $A\cup B$，即 $A\cup B=\{x\,|\,x\in A$或$x\in B\}$. 显然 $A\subseteq A\cup B$，$B\subseteq A\cup B$.

由 A 和 B 的公共元素组成的集合称为 A 与 B 的**交集**，记作 $A\cap B$，即 $A\cap B=$

$\{x|x\in A$且$x\in B\}$. 显然$A\cap B\subseteq A$，$A\cap B\subseteq B$.

类似地，设给定n个集合A_1，A_2，…，A_n. 由A_1，A_2，…，A_n的所有元素组成的集合称为A_1，A_2，…，A_n的并；而由A_1，A_2，…，A_n的公共元素组成的集合称为A_1，A_2，…，A_n的交，分别记为$A_1\cup A_2\cup\cdots\cup A_n$和$A_1\cap A_2\cap\cdots\cap A_n$.

设A，B是两个集合，由一切属于A但不属于B的元素组成的集合称为B在A中的**余集**，或称为A与B的**差**，记作$A-B$. 即$A-B=\{x|x\in A$且$x\notin B\}$. 例如，$\boldsymbol{R}-\boldsymbol{Q}$就是全体无理数的集合.

设A，B是两个集合，集合$A+B=\{a+b|a\in A，b\in B\}$称为$A$与$B$的**和集**. 要注意和集与集合的并是两个完全不同的概念.

设A，B是两个集合，集合$A\times B=\{(a,b)|a\in A，b\in B\}$称为$A$与$B$的**积**.

以上我们对集合进行了简单的讨论.

另一个常用的基本概念是映射，以下来讨论映射和它的一些性质.

定义 1.1.1 设A，B是两个非空集合，A到B的一个映射是指一个对应法则，通过这一法则，对于集合A中的每一个元素x，都有集合B中的一个唯一确定的元素y与之对应. 我们用字母f，g，σ，μ，…表示映射，用记号f：$A\to B$表示f是A到B的一个映射. 如果通过映射f，与A中元素x对应的B中元素是y，则记作f：$x\mapsto y$或$f(x)=y$. y叫做元素x在f下的**象**，x叫做y在f下的**原象**. A在f下的象的集合记作$f(A)=\{f(x)|x\in A\}$. 如果对于每一个$x\in A$，$f(x)$都已给出，那么映射f就被确定了.

【例 1.1.1】 $\mathbf{R}^{n\times n}$是全体n阶实矩阵A组成的集合，令$f(A)=\det A$，则$\forall A\in\mathbf{R}^{n\times n}$，通过法则$f$，在$\mathbf{R}$中有唯一确定的实数$x=\det A$与之对应，所以$f$：$\mathbf{R}^{n\times n}\to\mathbf{R}$是一个映射.

【例 1.1.2】 设$\mathbf{R}^+$是一切正实数的集合，$\forall x\in\mathbf{R}^+$，令$f(x)=\pm\sqrt{x}$与之对应. 由于$x>0$时，$f(x)$不能由x唯一确定，所以f不是$\mathbf{R}^+$到$\mathbf{R}$的映射.

【例 1.1.3】 设$P_n[x]$是所有次数不大于n的实系数多项式的集合，$\forall f(x)\in P_n[x]$，令$\sigma[f(x)]=f'(x)$，则σ是$P_n[x]$到自身的一个映射 [其中$f'(x)$表示$f(x)$对x求一阶导数].

某个集合A到自身的映射也称为A的一个**变换**.

【例 1.1.4】 设A是任意一个非空集合，$\forall x\in A$，令$f(x)=x$与之对应：f：$x\mapsto x$. 显然这是A到A的一个映射，这个映射称为A的**恒等映射**（或**恒等变换**），记作j_A.

关于两个集合间的映射有以下几点需要注意：

(1) A、B可以是相同的集合，也可以是不同的集合；

(2) 对于A中的每一个元素x，B中必有一个唯一确定的元素与之对应；

(3) 一般来说，B中的元素不一定都是A中元素的象；

(4) A中不同元素的象可能相同.

定义 1.1.2 设f，g都是A到B的映射，如果$\forall x\in A$，都有$f(x)=g(x)$，则称映射f与g是相等的，记作$f=g$.

定义 1.1.3 设f是A到B的一个映射. 如果$f(A)=B$，则称f是A到B上的一个映射，也称f是一个满射.

定义 1.1.4 设f是A到B的一个映射. 如果对于A中的任意两个元素x_1和x_2，当

$x_1 \neq x_2$ 时，有 $f(x_1) \neq f(x_2)$，则称 f 是 A 到 B 的一个单射.

定义 1.1.5 设 f 是 A 到 B 的一个映射，g 是 B 到 C 的一个映射，若对于每一个 $x \in A$，有 C 中的一个唯一确定的元素 $g[f(x)]$ 与之对应，则得到一个由 A 到 C 的映射 gf，称其为 f 与 g 的合成（或称为映射的积）.

事实 设给定映射 f：$A \to B$，g：$B \to C$，h：$C \to D$，则合成映射 $h(gf)$ 与 $(hg)f$ 都是 A 到 D 的映射. 并且有 $h(gf) = (hg)f$.

证 由定义 1.1.5 易见 $h(gf)$ 与 $(hg)f$ 都是 A 到 D 的映射.

设 $u = gf$，$v = hg$，则 $\forall x \in A$，有

$$hu(x) = h(u(x)) = h(g(f(x))), vf(x) = v(f(x)) = h(g(f(x)))$$

$\therefore hu = vf, h(gf) = (hg)f$.

一般的，映射的合成不满足交换律，即 $fg \neq gf$.

定义 1.1.6 如果映射 f：$A \to B$ 既是满射又是单射，即 f 满足以下条件：

(1) $f(A) = B$；

(2) $\forall x_1$，$x_2 \in A$，若 $f(x_1) = f(x_2)$，则 $x_1 = x_2$.

则称 f 是 A 到 B 的一个 1－1 映射.

特别地，一个有限集合 A 到自身的 1－1 映射称为 A 的一个**置换**. 显然，如果 A 是有限集合，则恒等映射 j_A 是一个置换，这个置换也称为**恒等置换**.

定理 1.1.1 设 f：$A \to B$ 是一个映射，则以下两个条件等价：

(1) f 是 1－1 映射；

(2) 存在映射 g：$B \to A$，使得 $gf = j_A$，$fg = j_B$.

并且当条件 (2) 成立时，映射 g 由映射 f 唯一确定.

证 若条件 (1) 成立，$\because f(A) = B$，$\therefore \forall y \in B$，有 $x \in A$，使得 $f(x) = y$；

又因为 f 是单射，所以这个 x 是由 y 唯一确定的. 规定 g：$y \mapsto x$.

如果 $f(x) = y$，则 g 是 B 到 A 的一个映射.

设 $x \in A$，而 $f(x) = y$. 则有 $gf(x) = g(f(x)) = g(y) = x$，$\therefore gf = j_A$.

设 $y \in B$，而 $f(x) = y$，那么 $g(y) = x$，$\therefore fg(y) = f(g(y)) = f(x) = y$，

$\therefore fg = j_B$.

反之，若条件 (2) 成立，设 $y \in B$，令 $g(y) = x \in A$. $\because fg = j_B$，

$\therefore f(x) = f(g(y)) = j_B(y) = y$. $\therefore f$ 是满射.

再设 x_1，$x_2 \in A$，而 $f(x_1) = f(x_2)$，

$\therefore x_1 = j_A(x_1) = gf(x_1) = gf(x_2) = j_A(x_2) = x_2$. $\therefore f$ 是单射.

所以 f 是 1－1 映射.

此外，当条件 (2) 成立时，令 g：$B \to A$，h：$B \to A$ 均具有性质

$gf = hf = j_A$，$fg = fh = j_B$，则有 $g = gj_B = g(fh) = (gf)h = j_A h = h$. 所以 g 由 f 唯一确定.

定义 1.1.7 满足定理 1.1.1 条件 (2) 的映射 g：$B \to A$ 叫做 f 的逆映射，记作 f^{-1}. 显然有 $f^{-1}f = j_A$，$ff^{-1} = j_B$，$(f^{-1})^{-1} = f$.

需要注意的是：一个映射不一定有逆映射.

§1.2 线性空间及其性质

在解析几何中，我们已经学习过平面与空间的向量，线性空间是解析几何中向量概念的一般化.

定义 1.2.1 设 P 是一个数域，P 中的元素用小写拉丁字母 a，b，c，…表示，令 $\mathbf{V}$ 是一个非空集合，$\mathbf{V}$ 中的元素用小写希腊字母 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$，$\boldsymbol{\gamma}$，…表示. 称 $\mathbf{V}$ 中的元素为向量，P 中的元素为纯量. 如果下列条件被满足，则称 $\mathbf{V}$ 是 P 上的一个线性空间（或向量空间）：

（1）在 $\mathbf{V}$ 中定义了一个“加法”. 对于 $\mathbf{V}$ 中任意两个向量 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$，有 $\mathbf{V}$ 中一个唯一确定的向量与它们对应，这个向量叫做 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的和，记作 $\boldsymbol{\alpha}+\boldsymbol{\beta}$.

（2）有一个“纯量乘法”. 对于 P 中每一个纯量 a 和 $\mathbf{V}$ 中每一个向量 $\boldsymbol{\alpha}$，有 $\mathbf{V}$ 中一个唯一确定的向量与它们对应，这个向量叫做 a 与 $\boldsymbol{\alpha}$ 的积，记作 $a\boldsymbol{\alpha}$.

（3）向量的加法和纯量的乘法满足下列运算律：

1）$\boldsymbol{\alpha}+\boldsymbol{\beta}=\boldsymbol{\beta}+\boldsymbol{\alpha}$；

2）$(\boldsymbol{\alpha}+\boldsymbol{\beta})+\boldsymbol{\gamma}=\boldsymbol{\alpha}+(\boldsymbol{\beta}+\boldsymbol{\gamma})$；

3）在 $\mathbf{V}$ 中存在一个零向量，记作 $\mathbf{0}$，对于 $\mathbf{V}$ 中每一个向量 $\boldsymbol{\alpha}$，都有 $\mathbf{0}+\boldsymbol{\alpha}=\boldsymbol{\alpha}$；

4）对于 $\mathbf{V}$ 中每一个向量 $\boldsymbol{\alpha}$，在 $\mathbf{V}$ 中存在一个向量 $\boldsymbol{\alpha}'$，使得 $\boldsymbol{\alpha}+\boldsymbol{\alpha}'=\mathbf{0}$，这样的 $\boldsymbol{\alpha}'$ 叫做 $\boldsymbol{\alpha}$ 的负向量，记作 $\boldsymbol{\alpha}'=-\boldsymbol{\alpha}$；

5）$a(\boldsymbol{\alpha}+\boldsymbol{\beta})=a\boldsymbol{\alpha}+a\boldsymbol{\beta}$；

6）$(a+b)\boldsymbol{\alpha}=a\boldsymbol{\alpha}+b\boldsymbol{\alpha}$；

7）$(ab)\boldsymbol{\alpha}=a(b\boldsymbol{\alpha})$；

8）$1\cdot\boldsymbol{\alpha}=\boldsymbol{\alpha}$.

其中 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$，$\boldsymbol{\gamma}$ 是 $\mathbf{V}$ 中的任意向量，a，b 是 P 中的纯量.

$\mathbf{V}$ 中所定义的加法与纯量乘法称为 $\mathbf{V}$ 中的线性运算，在不致发生混淆的情况下，将数域 P 上的线性空间 $\mathbf{V}$ 简称为线性空间 $\mathbf{V}$. 当 P 为实数域 $\mathbf{R}$，或复数域 $\mathbf{C}$ 时，分别称 $\mathbf{V}$ 为实线性空间，或复线性空间.

【例 1.2.1】 在解析几何里，二维平面或三维空间中的所有向量，对于向量的加法和实数与向量的乘法这两种线性运算构成实数域 $\mathbf{R}$ 上的线性空间. 分别记作 $\mathbf{R}^2$，$\mathbf{R}^3$.

【例 1.2.2】 分量属于数域 P 的全体 n 元数组 $(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 构成 P 上的一个线性空间，记作 P^n. 当 $P=\mathbf{C}$ 时，P^n 称为 n 元复线性空间，记作 $\mathbf{C}^n$；当 $P=\mathbf{R}$ 时，P^n 称为 n 元实线性空间，记作 $\mathbf{R}^n$.

【例 1.2.3】 数域 P 上一切 $m\times n$ 矩阵所构成的集合，对于矩阵的加法和数与矩阵的乘法构成 P 上的一个线性空间.

当 $P=\mathbf{C}$ 时，则此空间记作 $\mathbf{C}^{m\times n}$；当 $P=\mathbf{R}$ 时，则此空间记作 $\mathbf{R}^{m\times n}$.

特别地，P 上一切 1 行 n 列矩阵所成的集合和一切 n 行 1 列矩阵所成的集合分别构成 P 上的线性空间. 前者称为 P 上的行空间，后者称为 P 上的列空间.

【例 1.2.4】 给定 $A\in\mathbf{C}^{m\times n}$，记 $R(A)=\{y\mid y=Ax, x\in\mathbf{C}^n\}$，$N(A)=\{x\mid Ax=0, x\in\mathbf{C}^n\}$，按 $\mathbf{C}^n$ 中的加法和数乘运算，$R(A)$，$N(A)$ 都是 $\mathbf{C}$ 上的线性空间.

证 （1）设 $y_1, y_2\in R(A)$，则存在 x_1，$x_2\in\mathbf{C}^n$，使得 $y_1=Ax_1$，$y_2=Ax_2$，

$\because \mathbf{C}^n$ 是线性空间，$\therefore x_1+x_2\in\mathbf{C}^n$，$\therefore A(x_1+x_2)\in R(A)$

又$\because A(x_1+x_2)=Ax_1+Ax_2=y_1+y_2,\therefore y_1+y_2\in R(A)$.

(2) 同理，设 $k\in\mathbf{C}$，$\therefore kx_1\in\mathbf{C}^n$，则 $A(kx_1)\in R(A)$，

又$\because A(kx_1)=kAx_1=ky_1,\therefore ky_1\in R(A)$.

(3) 1)$y_1+y_2=Ax_1+Ax_2=A(x_1+x_2)=A(x_2+x_1)=Ax_2+Ax_1=y_2+y_1$;

2)$(y_1+y_2)+y_3=(Ax_1+Ax_2)+Ax_3=Ax_1+(Ax_2+Ax_3)=y_1+(y_2+y_3)$;

3)$\because \mathbf{0}\in\mathbf{C}^n,\therefore \mathbf{0}=A\mathbf{0}\in R(A)$;

4) 设 $x_1\in\mathbf{C}^n,\therefore -x_1\in\mathbf{C}^n,\therefore \mathbf{0}=Ax_1+A(-x_1)=A(x_1-x_1)\in R(A)$;

5) 设 $a\in\mathbf{C},a(y_1+y_2)=a(Ax_1+Ax_2)=aAx_1+aAx_2=ay_1+ay_2$;

6) 设 $a,b\in\mathbf{C},(a+b)y_1=(a+b)Ax_1=aAx_1+bAx_1=ay_1+by_1$;

7)$(ab)y_1=(ab)Ax_1=a(bAx_1)=a(by_1)$;

8)$1\cdot y_1=1\cdot Ax_1=Ax_1=y_1$.

所以 $R(A)$ 为 $\mathbf{C}$ 上的线性空间. 同理可证明 $N(A)$ 也是 $\mathbf{C}$ 上的线性空间.

【例 1.2.5】 复数域 $\mathbf{C}$ 可以视为实数域 $\mathbf{R}$ 上的线性空间.

事实上，二复数之和还是复数，一个实数与一个复数之积还是复数，同时定义 1.2.1 中的（3）也成立.

【例 1.2.6】 任意数域 P 总可以视为它自身上的线性空间.

【例 1.2.7】 全体实函数，按函数的加法和实数与函数的乘法构成 $\mathbf{R}$ 上的线性空间.

【例 1.2.8】 仅由 $\mathbf{C}$ 上线性空间 $\mathbf{V}$ 中的零元素 $\mathbf{0}$ 构成的单元素集合 $\mathbf{0}=\{\mathbf{0}|\mathbf{0}\in\mathbf{V}\}$，按 $\mathbf{V}$ 中的运算定义运算，则 $\mathbf{0}$ 是 $\mathbf{C}$ 上的一个线性空间，这个空间叫做零空间.

【例 1.2.9】 问当 $\boldsymbol{b}\neq 0$ 时，相容的线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的解的全体 $\boldsymbol{S}=\{\boldsymbol{x}\mid \boldsymbol{Ax}=\boldsymbol{b},\boldsymbol{x}\in\boldsymbol{C}^n\}$ 是否构成线性空间?

证 （反证法）若 $\boldsymbol{S}$ 是线性空间，则 $\forall x_1, x_2\in\boldsymbol{S}$，有 $x_1+x_2\in\boldsymbol{S}$.

但是 $\boldsymbol{A}(x_1+x_2)=\boldsymbol{A}x_1+\boldsymbol{A}x_2=2b\neq b$，$\therefore x_1+x_2\notin S$.

所以 $\boldsymbol{S}$ 不是线性空间.

【例 1.2.10】 设 $\{a_n\}$，$\{b_n\}$ 是两个收敛于 0 的实数无穷序列，则 $\lim\limits_{n\to\infty}(a_n+b_n)=\lim\limits_{n\to\infty}a_n+\lim\limits_{n\to\infty}b_n=0$；且 $\forall a\in\boldsymbol{R}$，有 $\lim\limits_{n\to\infty}aa_n=a\lim\limits_{n\to\infty}a_n=0$；

易证定义 1.2.1 中的（3）也成立.

所以，一切收敛于 0 的实序列对于如上定义的加法和数与序列的乘法构成 $\boldsymbol{R}$ 上的一个线性空间.

【例 1.2.11】 n 阶线性齐次常微分方程 $L[x]=\dfrac{\mathrm{d}^nx}{\mathrm{d}t^n}+a_1\dfrac{\mathrm{d}^{n-1}x}{\mathrm{d}t^{n-1}}+a_2\dfrac{\mathrm{d}^{n-2}x}{\mathrm{d}t^{n-2}}+\cdots+a_{n-1}\dfrac{\mathrm{d}x}{\mathrm{d}t}+a_nx=0$ 的解的全体 $\boldsymbol{S}=\{x(t)\,|\,L[x]=0\}$，对于普通函数的加法、数乘法构成 $\mathbf{C}$ 上的线性空间.

对于零向量与负向量，我们有以下的定理.

定理 1.2.1 在线性空间 $\mathbf{V}$ 中，零向量是唯一的；对于 $\mathbf{V}$ 中的任意向量 $\boldsymbol{\alpha}$，其负向量也是唯一的.

证 设 $\mathbf{0}$ 和 $\mathbf{0}'$ 都是 $\mathbf{V}$ 的零向量，则 $\forall\boldsymbol{\alpha}\in\mathbf{V}$，有 $\mathbf{0}+\boldsymbol{\alpha}=\boldsymbol{\alpha}$，$\boldsymbol{\alpha}+\mathbf{0}'=\boldsymbol{\alpha}$.

$\therefore \mathbf{0}=\mathbf{0}+\mathbf{0}'=\mathbf{0}'$.

设 $\boldsymbol{\alpha}'$，$\boldsymbol{\alpha}''$都是 $\boldsymbol{\alpha}$ 的负向量，则有 $\boldsymbol{\alpha}'+\boldsymbol{\alpha}=\mathbf{0}$，$\boldsymbol{\alpha}''+\boldsymbol{\alpha}=\mathbf{0}$.

所以 $\boldsymbol{\alpha}'=\boldsymbol{\alpha}'+\mathbf{0}=\boldsymbol{\alpha}'+(\boldsymbol{\alpha}+\boldsymbol{\alpha}'')=(\boldsymbol{\alpha}'+\boldsymbol{\alpha})+\boldsymbol{\alpha}''=\mathbf{0}+\boldsymbol{\alpha}''=\boldsymbol{\alpha}''$.

定义 1.2.2 在线性空间 $\mathbf{V}$ 中，两个向量 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的差为 $\boldsymbol{\alpha}+(-\boldsymbol{\beta})$，记作 $\boldsymbol{\alpha}-\boldsymbol{\beta}$.

定理 1.2.2 对任意向量 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$，$\boldsymbol{\gamma}\in\mathbf{V}$，对任意的数 $a\in P$，有

(1) $0\boldsymbol{\alpha}=\mathbf{0}$，$a\mathbf{0}=\mathbf{0}$;

(2) $a(-\boldsymbol{\alpha})=(-a)\boldsymbol{\alpha}=-a\boldsymbol{\alpha}$;

(3) $a\boldsymbol{\alpha}=\mathbf{0}\Rightarrow a=0$或 $\boldsymbol{\alpha}=\mathbf{0}$;

(4) $\boldsymbol{\alpha}+\boldsymbol{\beta}=\boldsymbol{\gamma}\Leftrightarrow\boldsymbol{\alpha}=\boldsymbol{\gamma}-\boldsymbol{\beta}$.

证 (1) $0\boldsymbol{\alpha}=0\boldsymbol{\alpha}+\mathbf{0}=0\boldsymbol{\alpha}+(0\boldsymbol{\alpha}-0\boldsymbol{\alpha})=(0\boldsymbol{\alpha}+0\boldsymbol{\alpha})-0\boldsymbol{\alpha}=(0+0)\boldsymbol{\alpha}-0\boldsymbol{\alpha}=0\boldsymbol{\alpha}-0\boldsymbol{\alpha}=\mathbf{0}$

同理可证 $a\mathbf{0}=\mathbf{0}$.

(2) $\because a\boldsymbol{\alpha}+a(-\boldsymbol{\alpha})=a(\boldsymbol{\alpha}+(-\boldsymbol{\alpha}))=a\mathbf{0}=\mathbf{0}$

$$a(-\boldsymbol{\alpha})=-a\boldsymbol{\alpha}$$

同理可证 $(-a)\boldsymbol{\alpha}=-a\boldsymbol{\alpha}$.

(3) 设 $a\boldsymbol{\alpha}=\mathbf{0}$，但 $a\neq0$，

则 $\alpha=1\boldsymbol{\alpha}=\left(\frac{1}{a}a\right)\boldsymbol{\alpha}=\frac{1}{a}(a\boldsymbol{\alpha})=\frac{1}{a}\mathbf{0}=\mathbf{0}$.

(4) $\because\ \boldsymbol{\alpha}+\boldsymbol{\beta}=\boldsymbol{\gamma}\Leftrightarrow\boldsymbol{\alpha}+\boldsymbol{\beta}+(-\boldsymbol{\beta})=\boldsymbol{\gamma}+(-\boldsymbol{\beta})\Leftrightarrow\boldsymbol{\alpha}+(\boldsymbol{\beta}-\boldsymbol{\beta})=\boldsymbol{\gamma}-\boldsymbol{\beta}$

$$\Leftrightarrow\boldsymbol{\alpha}+\mathbf{0}=\boldsymbol{\gamma}-\boldsymbol{\beta}\Leftrightarrow\boldsymbol{\alpha}=\boldsymbol{\gamma}-\boldsymbol{\beta}$$

由前面所述可见，对于给定的数域 P，分量取自 P 的 n 元数组 $P^n(n\in\mathbf{Z})$，在通常的运算下（在 P^n 中按分量相加）构成 P 上的一个线性空间. 特别是，$\mathbf{R}^n$，$\mathbf{C}^n$ 为本教材的基本线性空间. 具有实系数或复系数的（不超过某一指定次数的，或任意次数的）多项式集合，以及区间$[a,b]\subset\mathbf{R}$上的实（复）值连续函数，或任意函数的集合也是（$\mathbf{R}$ 上或 $\mathbf{C}$ 上）线性空间的例子.

当然，有限维空间 $\mathbf{R}^n$ 与由$[0,1]$上的实值连续函数组成的无限维线性空间之间，有着本质的区别.

§1.3 基、维数与坐标

线性空间中相关性概念与线性代数中向量组线性相关性概念类似.

设 V 是 P 上的一个线性空间，则有

(1) 线性组合：$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_m\in\mathbf{V}$，$\lambda_1$，$\lambda_2$，…，$\lambda_m\in P$

$$\lambda_1\boldsymbol{\alpha}_1+\lambda_2\boldsymbol{\alpha}_2+\cdots+\lambda_m\boldsymbol{\alpha}_m$$

称为元素组 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_m$ 的一个线性组合.

(2) 线性表示：$\mathbf{V}$ 中某个元素 $\boldsymbol{x}$ 可表示为其中某个元素组的线性组合，则称 $\boldsymbol{x}$ 可由该元素组线性表示.

(3) 线性相关性：如果存在一组不全为零的数 λ_1，λ_2，…，$\lambda_m\in P$，使得对于元素 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_m\in V$ 有

$$\lambda_1\boldsymbol{\alpha}_1+\lambda_2\boldsymbol{\alpha}_2+\cdots+\lambda_m\boldsymbol{\alpha}_m=\mathbf{0}$$

则称元素组 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_m$ 线性相关，否则称其线性无关.

【例 1.3.1】 讨论 $P_3[t]$ 中的多项式组，$f_1(t)=1+4t-2t^2+t^3$，$f_2(t)=-1+9t-3t^2+2t^3$，$f_3(t)=-5+6t+t^3$，$f_4(t)=5+7t-5t^2+2t^3$ 的线性相关性.

解 设 $k_i(i=1,2,3,4)$，使得 $k_1f_1(t)+k_2f_2(t)+k_3f_3(t)+k_4f_4(t)=0$，代入 $f_i(t)$ 整理后得 $(k_1-k_2-5k_3+5k_4)+(4k_1+9k_2+6k_3+7k_4)t+(-2k_1-3k_2-5k_4)t^2+(k_1+2k_2+k_3+2k_4)t^3=0$.

因为 1，t，t^2，t^3 线性无关，所以

$$\begin{cases} k_1-k_2-5k_3+5k_4=0 \\ 4k_1+9k_2+6k_3+7k_4=0 \\ -2k_1-3k_2-5k_4=0 \\ k_1+2k_2+k_3+2k_4=0 \end{cases}$$

因为此方程组的系数行列式 $D=0$，所以它有非零解，故 $f_1(t)$，$f_2(t)$，$f_3(t)$，$f_4(t)$ 线性相关.

线性相关性概念是个非常重要的概念，我们应用它来研究线性空间.

定义 1.3.1 设 $\boldsymbol{V}$ 是数域 P 上的一个线性空间，$\boldsymbol{V}$ 中满足以下两个条件的向量组 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 叫做 $\boldsymbol{V}$ 的一个基.

(1) $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 线性无关；

(2) $\boldsymbol{V}$ 中的每一个向量都可以由 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 线性表示.

其中 n 称为线性空间 $\boldsymbol{V}$ 的维数，记作 $\dim\boldsymbol{V}=n$. 维数为 n 的线性空间称为 n 维线性空间，记作 $\boldsymbol{V}^n$.

在线性代数中，我们曾经学习了极大线性无关组的概念. 实际上，线性空间的基就是线性空间的极大线性无关组.

【例 1.3.2】 在线性空间 $\mathbf{R}^2$ 中，任意两个不共线的非零向量都构成 $\mathbf{R}^2$ 的一个基；在线性空间 $\mathbf{R}^3$ 中，任意三个不共面的非零向量都构成 $\mathbf{R}^3$ 的一个基. 并且 $\dim\mathbf{R}^2=2$，$\dim\mathbf{R}^3=3$.

【例 1.3.3】 设 $\boldsymbol{C}^{m\times n}$ 是数域 $\boldsymbol{C}$ 上所有 $m\times n$ 矩阵构成的线性空间，考虑如下的 mn 个矩阵

$$\boldsymbol{E}_{ij}=\begin{pmatrix} & & & 0 & & & \\ & & & \vdots & & & \\ & & & 0 & & & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ & & & 0 & & & \\ & & & \vdots & & & \\ & & & 0 & & & \end{pmatrix}\begin{matrix} \\ \\ \\ (i) \\ \\ \\ \\ \end{matrix}$$

$$(j)$$

在 $\boldsymbol{E}_{ij}$ 中，除去第 i 行第 j 列位置上的元素是 1 外，其余的元素都是 0，$(i=1, 2, \cdots, m; j=1, 2, \cdots, n)$. 由定义 1.3.1 可见 $\{\boldsymbol{E}_{ij}\}$ 构成所有 $m\times n$ 矩阵组成的线性空间 $\boldsymbol{C}^{m\times n}$ 的一个基，并且 $\dim\boldsymbol{C}^{m\times n}=m\times n$. 当然对于线性空间 $\boldsymbol{R}^{m\times n}$ 而言，也有 $\dim\boldsymbol{R}^{m\times n}=m\times n$. 显然，$\boldsymbol{R}^{m\times n}$，$\boldsymbol{C}^{m\times n}$ 有一个相同的基 $\{\boldsymbol{E}_{ij}\}$.

【例 1.3.4】 零空间的维数是零.

线性空间不一定是有限维的，如 $\mathbf{C}[a,b]$ 就是一个无限维的线性空间. 在矩阵论中只考虑有限维情况，对于无限维情况有兴趣的可阅读泛函分析及其他书籍的相关内容.

定理 1.3.1 设$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \cdots, \boldsymbol{\alpha}_n\}$是数域$P$上的线性空间$\mathbf{V}^n$的一个基，则$\mathbf{V}^n$中的每一个向量可以唯一地被表示为基向量$\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$的线性组合.

证 设$\forall \boldsymbol{\alpha} \in \mathbf{V}^n$，$\boldsymbol{\alpha}=x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_n\boldsymbol{\alpha}_n$

$$\boldsymbol{\alpha} = y_1\boldsymbol{\alpha}_1 + y_2\boldsymbol{\alpha}_2 + \cdots + y_n\boldsymbol{\alpha}_n$$

两式相减有$(x_1-y_1)\boldsymbol{\alpha}_1+(x_2-y_2)\boldsymbol{\alpha}_2+\cdots+(x_n-y_n)\boldsymbol{\alpha}_n=\mathbf{0}$

$\because \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$线性无关，$\therefore x_1=y_1, x_2=y_2, \cdots, x_n=y_n$.

反之，任给一个有序数组$x_1, x_2, \cdots, x_n$，总有唯一的元素$\boldsymbol{\alpha} \in \mathbf{V}^n$可以由$\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$线性表示为$\boldsymbol{\alpha}=x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_n\boldsymbol{\alpha}_n \in \mathbf{V}^n$（其中$x_1, x_2, \cdots, x_n \in P$）.

由此可知，如果$\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$是$\mathbf{V}^n$的一个基，则$\mathbf{V}^n$中元素的全体可以表示为

$$\mathbf{V}^n = \{\boldsymbol{\alpha} \mid \boldsymbol{\alpha} = x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n, x_1, x_2, \cdots, x_n \in P\}$$

也就是说$\mathbf{V}^n$中的向量$\boldsymbol{\alpha}$与有序数组$x_1, x_2, \cdots, x_n$之间构成一一对应关系.

定义 1.3.2 设$\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$是线性空间$\mathbf{V}^n$的一个基，$\forall \boldsymbol{\alpha} \in \mathbf{V}^n$，有且仅有一个有序数组$x_1, x_2, \cdots, x_n$，使得$\boldsymbol{\alpha}=x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_n\boldsymbol{\alpha}_n$，我们称$x_1, x_2, \cdots, x_n$为元素$\boldsymbol{\alpha}$在基$\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$下的坐标，记作$(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$.

【例 1.3.5】 取定$\mathbf{R}^3$中三个不共面的向量$\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$，$\forall \xi \in \mathbf{R}^3$，$\xi$可以唯一地表示为$\xi=x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+x_3\boldsymbol{\alpha}_3$. 因此，向量$\xi$关于基$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3\}$的坐标为$(x_1, x_2, x_3)^{\mathrm{T}}$.

【例 1.3.6】 在$\mathbf{R}^n$中如下的n个向量：$\varepsilon_i = (0,\cdots,0,\underset{(i)}{1},0,\cdots,0)^{\mathrm{T}}$，$i=1, 2, \cdots, n$作成$\mathbf{R}^n$的一个基，此基称为$\mathbf{R}^n$的标准基.

$\because \forall \boldsymbol{\alpha} = (a_1, a_2, \cdots, a_n)^{\mathrm{T}} \in \mathbf{R}^n$，有

$$\boldsymbol{\alpha} = a_1\varepsilon_1 + a_2\varepsilon_2 + \cdots + a_n\varepsilon_n$$

$\therefore \boldsymbol{\alpha}$关于基$\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n\}$的坐标为$(a_1, a_2, \cdots, a_n)^{\mathrm{T}}$.

【例 1.3.7】 在$\mathbf{R}^n$中如下的n个向量

$$\begin{aligned} \varepsilon'_1 &= (1,1,1,\cdots,1)^{\mathrm{T}}, \\ \varepsilon'_2 &= (0,1,1,\cdots,1)^{\mathrm{T}}, \\ \varepsilon'_3 &= (0,0,1,\cdots,1)^{\mathrm{T}}, \\ &\vdots \\ \varepsilon'_n &= (0,0,\cdots,0,1)^{\mathrm{T}} \end{aligned}$$

也是$\mathbf{R}^n$的一个基，因为$\forall \boldsymbol{\alpha} = (a_1, a_2, \cdots, a_n)^{\mathrm{T}} \in \mathbf{R}^n$，有

$$\boldsymbol{\alpha} = a_1\varepsilon'_1 + (a_2 - a_1)\varepsilon'_2 + \cdots + (a_n - a_{n-1})\varepsilon'_n$$

$\therefore \boldsymbol{\alpha}$关于基$\{\varepsilon'_1, \varepsilon'_2, \cdots, \varepsilon'_n\}$的坐标为$(a_1, a_2-a_1, \cdots, a_n-a_{n-1})^{\mathrm{T}}$.

【例 1.3.8】 求线性空间$P_n[x]$的基、维数以及向量p的坐标.

在$P_n[x]$中，它的一个基为$p_1=1$，$p_2=x$，$p_3=x^2$，$\cdots$，$p_n=x^{n-1}$，$p_{n+1}=x^n$；

$\therefore \dim P_n[x] = n+1$.

任意一个次数不大于n的多项式$p=a_0+a_1x+a_2x^2+\cdots+a_{n-1}x^{n-1}+a_nx^n$可以表示为

$$p = a_0p_1 + a_1p_2 + a_2p_3 + \cdots + a_{n-1}p_n + a_np_{n+1}$$

$\therefore p$在基$\{p_1, p_2, \cdots, p_{n+1}\}$下的坐标为$p = (a_0, a_1, a_2, \cdots, a_n)^{\mathrm{T}}$.

如果在$P_n[x]$中另取一个基$\{p'_1, p'_2, \cdots, p'_{n+1}\}$，其中$p'_1=1$，$p'_2=x-a$，$\cdots$，$p'_{n+1}=(x-a)^n$，则由$p$在$x=a$点的 Taylor 多项式$p = p(a) + p'(a)(x-a) + \cdots + \dfrac{p^{(n)}(a)}{n!}$

$(x-a)^n$ 可知 p 在基 $\{p'_1, p'_2, \cdots, p'_{n+1}\}$ 下的坐标为 $p=\left(p(a),p'(a),\cdots,\frac{p^{(n)}(a)}{n!}\right)^{\mathrm{T}}$.

【例 1.3.9】 线性空间 $\mathbf{C}^{m\times n}$ 的一个基为

$$\boldsymbol{E}_{ij}=\begin{pmatrix} & & & 0 & & & \\ & & & \vdots & & & \\ & & & 0 & & & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ & & & 0 & & & \\ & & & \vdots & & & \\ & & & 0 & & & \end{pmatrix}\begin{matrix} \\ \\ \\ (i) \\ \\ \\ \\ \end{matrix} \quad (i=1,2,\cdots,m; j=1,2,\cdots,n)$$
$$\qquad\qquad\qquad\quad (j)$$

$\forall \boldsymbol{M}\in\mathbf{C}^{m\times n}$，有

$$\boldsymbol{M}=a_{11}\boldsymbol{E}_{11}+a_{12}\boldsymbol{E}_{12}+\cdots+a_{1n}\boldsymbol{E}_{1n}+a_{21}\boldsymbol{E}_{21}+\cdots+a_{mn}\boldsymbol{E}_{mn}=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}\boldsymbol{E}_{ij}$$

$\therefore \dim\mathbf{C}^{m\times n}=mn$；$\boldsymbol{M}=(a_{ij})$ 在基 $\{\boldsymbol{E}_{ij}\}$ $(i=1, 2, \cdots, m; j=1, 2, \cdots, n)$ 下的坐标为

$$(a_{11},a_{12},\cdots,a_{1n},a_{21},\cdots,a_{mn})^{\mathrm{T}}$$

设 n 维线性空间 $\boldsymbol{V}^n$ 的向量 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$ 关于基 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 的坐标分别为 $(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 和 $(y_1, y_2, \cdots, y_n)^{\mathrm{T}}$，$\boldsymbol{\alpha}=x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_n\boldsymbol{\alpha}_n$，$\boldsymbol{\beta}=y_1\boldsymbol{\alpha}_1+y_2\boldsymbol{\alpha}_2+\cdots+y_n\boldsymbol{\alpha}_n$，那么 $\boldsymbol{\alpha}+\boldsymbol{\beta}=(x_1+y_1)\boldsymbol{\alpha}_1+(x_2+y_2)\boldsymbol{\alpha}_2+\cdots+(x_n+y_n)\boldsymbol{\alpha}_n$.

设 $a\in P$，那么 $a\boldsymbol{\alpha}=(ax_1)\boldsymbol{\alpha}_1+(ax_2)\boldsymbol{\alpha}_2+\cdots+(ax_n)\boldsymbol{\alpha}_n$.

所以有如下定理.

定理 1.3.2 设 $\boldsymbol{V}$ 是数域 P 上的一个 n 维线性空间，$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 是 $\boldsymbol{V}$ 的一个基，$\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\boldsymbol{V}$，它们关于基 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 的坐标分别是 $(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 和 $(y_1, y_2, \cdots, y_n)^{\mathrm{T}}$，那么，$\boldsymbol{\alpha}+\boldsymbol{\beta}$ 关于这个基的坐标就是 $(x_1+y_1, x_2+y_2, \cdots, x_n+y_n)^{\mathrm{T}}$.

又设 $a\in P$，那么，$a\boldsymbol{\alpha}$ 关于这个基的坐标就是 $(ax_1, ax_2, \cdots, ax_n)^{\mathrm{T}}$.

一般来说，线性空间及其元素是抽象的对象，不同空间的元素完全可以具有千差万别的类别及性质. 但坐标表示却把它们统一了起来，坐标表示把这种差别留给了基和基元素，由坐标所组成的新向量仅由数域中的数表示出来. 更进一步，原本抽象的“加法”及“数乘”经过坐标表示就演化为向量加法及数对向量的数乘.

【例 1.3.10】 求 $\begin{pmatrix}1 & 1\\ 2 & -1\end{pmatrix}$，$\begin{pmatrix}2 & 4\\ 4 & -2\end{pmatrix}$，$\begin{pmatrix}3 & 4\\ 6 & -3\end{pmatrix}$，$\begin{pmatrix}-1 & 3\\ 4 & 4\end{pmatrix}$ 的极大线性无关组.

解 向量组 $\begin{pmatrix}1 & 1\\ 2 & -1\end{pmatrix}$，$\begin{pmatrix}2 & 4\\ 4 & -2\end{pmatrix}$，$\begin{pmatrix}3 & 4\\ 6 & -3\end{pmatrix}$，$\begin{pmatrix}-1 & 3\\ 4 & 4\end{pmatrix}$ 在自然基 $\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}$，$\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}$，$\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}$，$\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}$ 下的坐标分别是

$$\boldsymbol{\alpha}_1=(1,1,2,-1)^{\mathrm{T}},\boldsymbol{\alpha}_2=(2,4,4,-2)^{\mathrm{T}},\boldsymbol{\alpha}_3=(3,4,6,-3)^{\mathrm{T}},\boldsymbol{\alpha}_4=(-1,3,4,4)^{\mathrm{T}}$$

$$\begin{pmatrix}1 & 2 & 3 & -1\\ 1 & 4 & 4 & 3\\ 2 & 4 & 6 & 4\\ -1 & -2 & -3 & 4\end{pmatrix}\sim\begin{pmatrix}1 & 2 & 3 & -1\\ 0 & 2 & 1 & 4\\ 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0\end{pmatrix}$$

因此 $\begin{pmatrix}1 & 1\\2 & -1\end{pmatrix}$，$\begin{pmatrix}2 & 4\\4 & -2\end{pmatrix}$，$\begin{pmatrix}3 & 4\\6 & -3\end{pmatrix}$，$\begin{pmatrix}-1 & 3\\4 & 4\end{pmatrix}$ 的极大线性无关组为 $\begin{pmatrix}1 & 1\\2 & -1\end{pmatrix}$，$\begin{pmatrix}2 & 4\\4 & -2\end{pmatrix}$，$\begin{pmatrix}-1 & 3\\4 & 4\end{pmatrix}$.

【例 1.3.11】 求 $P_3[t]$ 中的多项式组，$f_1(t)=1+4t-2t^2+t^3$，$f_2(t)=-1+9t-3t^2+2t^3$，$f_3(t)=-5+6t+t^3$，$f_4(t)=5+7t-5t^2+2t^3$ 的秩和一个极大无关组.

解 取 $P_3[t]$ 的基 $1，t，t^2，t^3$. 以 $f_1(t),f_2(t),f_3(t),f_4(t)$ 在此基下的坐标为列向量构造矩阵 $\boldsymbol{A}$. 并且对 $\boldsymbol{A}$ 作初等行变换

$$\boldsymbol{A}=\begin{pmatrix}1 & -1 & -5 & 5\\4 & 9 & 6 & 7\\-2 & -3 & 0 & -5\\1 & 2 & 1 & 2\end{pmatrix}\xrightarrow[\substack{r_3+2r_1\\r_4-r_1}]{r_2-4r_1}\begin{pmatrix}1 & -1 & -5 & 5\\0 & 13 & 26 & -13\\0 & -5 & -10 & 5\\0 & 23 & 16 & -3\end{pmatrix}\xrightarrow[\substack{r_3+r_2\\r_4-3r_2}]{r_2\times\frac{1}{13}}\begin{pmatrix}1 & -1 & -5 & 5\\0 & 1 & 2 & -1\\0 & 0 & 0 & 0\\0 & 0 & 0 & 0\end{pmatrix}$$

可知 $\text{rank}\boldsymbol{A}=2$，且 $\boldsymbol{A}$ 的第 1、2 个列向量是 $\boldsymbol{A}$ 的列向量组的一个极大无关组；所以多项式组 $f_1(t),f_2(t),f_3(t),f_4(t)$ 的秩为 2，且 $f_1(t),f_2(t)$ 是一个极大无关组.

一个向量的坐标依赖于基的选取，对于线性空间 V 的两个不同的基来说，同一个向量的坐标一般是不相同的.

以下我们来讨论，一个向量关于不同的基的坐标之间的关系.

设 $\{\boldsymbol{\alpha}_1，\boldsymbol{\alpha}_2，\cdots，\boldsymbol{\alpha}_n\}$ 和 $\{\boldsymbol{\beta}_1，\boldsymbol{\beta}_2，\cdots，\boldsymbol{\beta}_n\}$ 是 V^n 的两个基，由此可知 $\boldsymbol{\beta}_j$ 作为 V^n 中的元素，可以表示为

$$\begin{aligned}\boldsymbol{\beta}_1&=a_{11}\boldsymbol{\alpha}_1+a_{21}\boldsymbol{\alpha}_2+\cdots+a_{n1}\boldsymbol{\alpha}_n\\\boldsymbol{\beta}_2&=a_{12}\boldsymbol{\alpha}_1+a_{22}\boldsymbol{\alpha}_2+\cdots+a_{n2}\boldsymbol{\alpha}_n\\&\vdots\\\boldsymbol{\beta}_n&=a_{1n}\boldsymbol{\alpha}_1+a_{2n}\boldsymbol{\alpha}_2+\cdots+a_{nn}\boldsymbol{\alpha}_n\end{aligned}\tag{1.3.1}$$

式中，$(a_{1j}，a_{2j}，\cdots，a_{nj})^{\mathrm{T}}$ 是 $\boldsymbol{\beta}_j$ 关于基 $\{\boldsymbol{\alpha}_1，\boldsymbol{\alpha}_2，\cdots，\boldsymbol{\alpha}_n\}$ 的坐标（$j=1，2，\cdots，n$）.

以这 n 个坐标为列，作一个 n 阶矩阵

$$\boldsymbol{T}=\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\a_{21} & a_{22} & \cdots & a_{2n}\\\vdots & \vdots & \ddots & \vdots\\a_{n1} & a_{n2} & \cdots & a_{nn}\end{pmatrix}$$

式中，矩阵 $\boldsymbol{T}$ 叫做由基 $\{\boldsymbol{\alpha}_1，\boldsymbol{\alpha}_2，\cdots，\boldsymbol{\alpha}_n\}$ 到基 $\{\boldsymbol{\beta}_1，\boldsymbol{\beta}_2，\cdots，\boldsymbol{\beta}_n\}$ 的过渡矩阵.

式（1.3.1）可以记为

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{T}\tag{1.3.2}$$

设 $\boldsymbol{\alpha}\in V$，$\boldsymbol{\alpha}$ 关于基 $\{\boldsymbol{\alpha}_1，\boldsymbol{\alpha}_2，\cdots，\boldsymbol{\alpha}_n\}$ 的坐标为 $(x_1，x_2，\cdots，x_n)^{\mathrm{T}}$，关于基 $\{\boldsymbol{\beta}_1，\boldsymbol{\beta}_2，\cdots，\boldsymbol{\beta}_n\}$ 的坐标为 $(y_1，y_2，\cdots，y_n)^{\mathrm{T}}$. 于是有

$$\boldsymbol{\alpha}=\sum_{i=1}^{n}x_i\boldsymbol{\alpha}_i=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\tag{1.3.3}$$

以及

$$\boldsymbol{\alpha}=\sum_{i=1}^{n}y_i\boldsymbol{\beta}_i=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix} \tag{1.3.4}$$

将式（1.3.2）代入式（1.3.4）有

$$\boldsymbol{\alpha}=((\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{T})\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\left[\boldsymbol{T}\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}\right] \tag{1.3.5}$$

由式（1.3.5）可见 $\boldsymbol{\alpha}$ 关于基$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$的坐标为$\boldsymbol{T}\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}$

但是我们知道向量 $\boldsymbol{\alpha}$ 关于一个基$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$的坐标是唯一的，所以

$$\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\boldsymbol{T}\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix} \tag{1.3.6}$$

于是有以下定理.

定理 1.3.3 设 $\boldsymbol{V}$ 是数域 P 上的一个 n 维线性空间，$\boldsymbol{T}$ 是由基$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$到基$\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n\}$的过渡矩阵，那么 V 中向量 $\boldsymbol{\alpha}$ 关于基$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$的坐标 $(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 与关于基$\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n\}$的坐标 $(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$ 由等式（1.3.6）联系着.

如果：$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$，$\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n\}$，$\{\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_n\}$都是线性空间 $\boldsymbol{V}^n$ 的基；由基$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$到基$\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n\}$的过渡矩阵为 $\boldsymbol{A}=(a_{ij})$；由基$\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n\}$，$\{\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_n\}$的过渡矩阵为$\boldsymbol{B}=(b_{ij})$. 则由以上的讨论可知

$$\boldsymbol{\beta}_k=\sum_{i=1}^{n}a_{ik}\boldsymbol{\alpha}_i,\quad k=1,2,\cdots,n \tag{1.3.7}$$

$$\boldsymbol{\gamma}_j=\sum_{k=1}^{n}b_{kj}\boldsymbol{\beta}_k,\quad j=1,2,\cdots,n \tag{1.3.8}$$

将式（1.3.7）代入式（1.3.8）有

$$\boldsymbol{\gamma}_j=\sum_{k=1}^{n}b_{kj}\sum_{i=1}^{n}a_{ik}\boldsymbol{\alpha}_i=\sum_{i=1}^{n}\Big(\sum_{k=1}^{n}a_{ik}b_{kj}\Big)\boldsymbol{\alpha}_i=\sum_{i=1}^{n}c_{ij}\boldsymbol{\alpha}_i \tag{1.3.9}$$

其中

$$c_{ij}=\sum_{k=1}^{n}a_{ik}b_{kj}. \tag{1.3.10}$$

令 $\boldsymbol{C}=(c_{ij})$，根据以上的讨论可知：由基$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$到基$\{\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_n\}$的过渡矩阵为$\boldsymbol{C}=(c_{ij})$；同时由式（1.3.10）亦可看出 $\boldsymbol{C}=\boldsymbol{AB}$.

∵由式（1.3.7）和式（1.3.8）可知

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A}$$

$$(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_n)=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\boldsymbol{B}$$

$\therefore\ (\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_n)=((\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A})\boldsymbol{B}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{AB}.$

由此可见，我们所熟悉的矩阵乘法的定义正反映了这种线性代入过程.

过渡矩阵反映了线性空间的不同的基之间的关系，这是一个很重要的概念，现在进一步讨论过渡矩阵的一些性质.

如果由基$\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n\}$到基$\{\boldsymbol{\beta}_1,\ \boldsymbol{\beta}_2,\ \cdots,\ \boldsymbol{\beta}_n\}$的过渡矩阵为$\boldsymbol{A}$

则有

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A}$$

再设基$\{\boldsymbol{\beta}_1,\ \boldsymbol{\beta}_2,\ \cdots,\ \boldsymbol{\beta}_n\}$到基$\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n\}$的过渡矩阵为$\boldsymbol{B}$，则有

$$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\boldsymbol{B}$$

比较上面两式有

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\boldsymbol{BA}$$
$$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{AB}$$

$\because\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n\}$，$\{\boldsymbol{\beta}_1,\ \boldsymbol{\beta}_2,\ \cdots,\ \boldsymbol{\beta}_n\}$都是基，则$\boldsymbol{AB}=\boldsymbol{BA}=\boldsymbol{E}$，

$\therefore$矩阵$\boldsymbol{A}$可逆，并且$\boldsymbol{B}=\boldsymbol{A}^{-1}$.

反之，设$\boldsymbol{A}=(a_{ij})$为任意一个可逆矩阵，任意取定$\boldsymbol{V}^n$的一个基$\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n\}$，并取$\boldsymbol{\beta}_j=\sum\limits_{i=1}^{n}a_{ij}\boldsymbol{\alpha}_i,j=1,2,\cdots,n$

则有

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A}$$

$$\therefore\quad(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\boldsymbol{A}^{-1}.$$

由此可知$\{\boldsymbol{\beta}_1,\ \boldsymbol{\beta}_2,\ \cdots,\ \boldsymbol{\beta}_n\}$也是$\boldsymbol{V}^n$的一个基.

因此有以下定理.

定理 1.3.4 设$\boldsymbol{A}$是n维线性空间$\boldsymbol{V}^n$中由基$\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n\}$到基$\{\boldsymbol{\beta}_1,\ \boldsymbol{\beta}_2,\ \cdots,\ \boldsymbol{\beta}_n\}$的过渡矩阵，那么$\boldsymbol{A}$是一个可逆矩阵；反之，任意一个$n$阶可逆矩阵$\boldsymbol{A}$都可以作为$\boldsymbol{V}^n$中一个基到另一个基的过渡矩阵. 如果由基$\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n\}$到基$\{\boldsymbol{\beta}_1,\ \boldsymbol{\beta}_2,\ \cdots,\ \boldsymbol{\beta}_n\}$的过渡矩阵为$\boldsymbol{A}$，那么，由基$\{\boldsymbol{\beta}_1,\ \boldsymbol{\beta}_2,\ \cdots,\ \boldsymbol{\beta}_n\}$到基$\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n\}$的过渡矩阵就是$\boldsymbol{A}^{-1}$.

【例 1.3.12】 在$\mathbf{R}^3$中，$\boldsymbol{\alpha}_1=(-3,\ 2,\ -1)$，$\boldsymbol{\alpha}_2=(0,\ -1,\ 2)$，$\boldsymbol{\alpha}_3=(-1,\ 2,\ -1)$，试证明$\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \boldsymbol{\alpha}_3\}$是$\mathbf{R}^3$的一个基，并求出$\boldsymbol{\alpha}=(3,\ 12,\ 7)$关于此基的坐标.

解 取$\mathbf{R}^3$中的标准基$\boldsymbol{\varepsilon}_1=(1,\ 0,\ 0)$，$\boldsymbol{\varepsilon}_2=(0,\ 1,\ 0)$，$\boldsymbol{\varepsilon}_3=(0,\ 0,\ 1)$，令

$$\boldsymbol{A}=\begin{pmatrix}-3 & 0 & -1\\ 2 & -1 & 2\\ -1 & 2 & -1\end{pmatrix},\ \therefore(\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \boldsymbol{\alpha}_3)=(\boldsymbol{\varepsilon}_1,\ \boldsymbol{\varepsilon}_2,\ \boldsymbol{\varepsilon}_3)\boldsymbol{A}$$

$\because\det\boldsymbol{A}=4\neq0$，$\therefore\boldsymbol{A}$可逆.

则$\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \boldsymbol{\alpha}_3\}$是$\mathbf{R}^3$的一个基，且$\boldsymbol{A}$为标准基到基$\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \boldsymbol{\alpha}_3\}$的过渡矩阵.

又$\because\boldsymbol{\alpha}=(3,\ 12,\ 7)$是$\boldsymbol{\alpha}$关于标准基$\{\boldsymbol{\varepsilon}_1,\ \boldsymbol{\varepsilon}_2,\ \boldsymbol{\varepsilon}_3\}$的坐标.

设$\boldsymbol{\alpha}$关于基$\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \boldsymbol{\alpha}_3\}$的坐标为$(x_1,\ x_2,\ x_3)$.

$$\because\begin{pmatrix}3\\12\\7\end{pmatrix}=\boldsymbol{A}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix},\ \therefore\boldsymbol{A}^{-1}\begin{pmatrix}3\\12\\7\end{pmatrix}=\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix},$$

则
$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=A^{-1}\begin{pmatrix}3\\12\\7\end{pmatrix}=\begin{pmatrix}-\frac{20}{3}\\\frac{46}{3}\\17\end{pmatrix}.$$

【例 1.3.13】 在 $\mathbf{R}^3$ 中有下列两组向量

$$\boldsymbol{\alpha}_1=(-3,1,-2),\boldsymbol{\alpha}_2=(1,-1,1),\boldsymbol{\alpha}_3=(2,3,-1)$$
$$\boldsymbol{\beta}_1=(1,1,1),\boldsymbol{\beta}_2=(1,2,3),\boldsymbol{\beta}_3=(2,0,1)$$

容易证明$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3\}$，$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3\}$都是 $\mathbf{R}^3$ 的基，试求由$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3\}$到$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3\}$的过渡矩阵.

∵由 $\mathbf{R}^3$ 的标准基到这两个基的过渡矩阵分别为

$$A=\begin{pmatrix}-3&1&2\\1&-1&3\\-2&1&-1\end{pmatrix},\quad B=\begin{pmatrix}1&1&2\\1&2&0\\1&3&1\end{pmatrix}$$

亦即$(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)=(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3)A$，$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)=(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3)B$.

∴ $(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3)A^{-1}B$

所求过渡矩阵为
$$A^{-1}B=\begin{pmatrix}-6&-19&-1\\-13&-42&-1\\-2&-7&0\end{pmatrix}.$$

【例 1.3.14】 设 $\boldsymbol{\alpha}\in\mathbf{R}^4$，$\boldsymbol{\alpha}$ 在基$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4\}$下的表达式为 $\boldsymbol{\alpha}=\boldsymbol{\alpha}_1-2\boldsymbol{\alpha}_2+3\boldsymbol{\alpha}_3+\boldsymbol{\alpha}_4$，求 $\boldsymbol{\alpha}$ 在基$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4\}$下的坐标. 此处

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\alpha}_4)\begin{pmatrix}1&0&0&0\\3&1&0&0\\-5&2&1&0\\7&-3&2&1\end{pmatrix}$$

解 设 $P=\begin{pmatrix}1&0&0&0\\3&1&0&0\\-5&2&1&0\\7&-3&2&1\end{pmatrix}$，$\therefore P^{-1}=\begin{pmatrix}1&0&0&0\\-3&1&0&0\\11&-2&1&0\\-38&7&-2&1\end{pmatrix}$

∵$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4)=(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4)P$ ∴ $\begin{pmatrix}y_1\\y_2\\y_3\\y_4\end{pmatrix}=P^{-1}\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}$

式中：$\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}$，$\begin{pmatrix}y_1\\y_2\\y_3\\y_4\end{pmatrix}$分别为 $\boldsymbol{\alpha}$ 在基$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4\}$与基$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4\}$下的坐标.

$$\therefore\begin{pmatrix}y_1\\y_2\\y_3\\y_4\end{pmatrix}=\begin{pmatrix}1&0&0&0\\-3&1&0&0\\11&-2&1&0\\-38&7&-2&1\end{pmatrix}\begin{pmatrix}1\\-2\\3\\1\end{pmatrix}=\begin{pmatrix}1\\-5\\18\\-57\end{pmatrix}$$

则

$$\boldsymbol{\alpha}=\boldsymbol{\beta}_1-5\boldsymbol{\beta}_2+18\boldsymbol{\beta}_3-57\boldsymbol{\beta}_4$$

【例 1.3.15】 设 $\mathbf{R}^{2\times2}$的两组基为：

①$\boldsymbol{A}_1=\begin{pmatrix}1&0\\0&0\end{pmatrix}$，$\boldsymbol{A}_2=\begin{pmatrix}1&1\\0&0\end{pmatrix}$，$\boldsymbol{A}_3=\begin{pmatrix}1&1\\1&0\end{pmatrix}$，$\boldsymbol{A}_4=\begin{pmatrix}1&1\\1&1\end{pmatrix}$；

②$\boldsymbol{B}_1=\begin{pmatrix}1&0\\1&1\end{pmatrix}$，$\boldsymbol{B}_2=\begin{pmatrix}0&1\\1&1\end{pmatrix}$，$\boldsymbol{B}_3=\begin{pmatrix}1&1\\1&0\end{pmatrix}$，$\boldsymbol{B}_4=\begin{pmatrix}1&1\\0&1\end{pmatrix}$.

试求：(1) 由基①到基②的过渡矩阵；

(2) 求在基①与基②下有相同坐标的矩阵.

解 (1) 取 $\mathbf{R}^{2\times2}$ 的基 $\boldsymbol{E}_{11}$，$\boldsymbol{E}_{12}$，$\boldsymbol{E}_{21}$，$\boldsymbol{E}_{22}$，则有 $(\boldsymbol{A}_1,\boldsymbol{A}_2,\boldsymbol{A}_3,\boldsymbol{A}_4)=(\boldsymbol{E}_{11},\boldsymbol{E}_{12},\boldsymbol{E}_{21},\boldsymbol{E}_{22})\boldsymbol{C}_1$，$(\boldsymbol{B}_1,\boldsymbol{B}_2,\boldsymbol{B}_3,\boldsymbol{B}_4)=(\boldsymbol{E}_{11},\boldsymbol{E}_{12},\boldsymbol{E}_{21},\boldsymbol{E}_{22})\boldsymbol{C}_2$

其中

$$\boldsymbol{C}_1=\begin{pmatrix}1&1&1&1\\0&1&1&1\\0&0&1&1\\0&0&0&1\end{pmatrix},\quad \boldsymbol{C}_2=\begin{pmatrix}1&0&1&1\\0&1&1&1\\1&1&1&0\\1&1&0&1\end{pmatrix}$$

于是 $(\boldsymbol{B}_1,\boldsymbol{B}_2,\boldsymbol{B}_3,\boldsymbol{B}_4)=(\boldsymbol{A}_1,\boldsymbol{A}_2,\boldsymbol{A}_3,\boldsymbol{A}_4)\boldsymbol{C}_1^{-1}\boldsymbol{C}_2=(\boldsymbol{A}_1,\boldsymbol{A}_2,\boldsymbol{A}_3,\boldsymbol{A}_4)\boldsymbol{C}$

所以由基①到基②的过渡矩阵为 $\boldsymbol{C}=\boldsymbol{C}_1^{-1}\boldsymbol{C}_2=\begin{pmatrix}1&-1&0&0\\-1&0&0&1\\0&0&1&-1\\1&1&0&1\end{pmatrix}$.

(2) 设 $\boldsymbol{A}\in\mathbf{R}^{2\times2}$，在基①与基②下坐标均为 $x=(x_1,x_2,x_3,x_4)^{\mathrm{T}}$，则由坐标变换公式 $x=\boldsymbol{C}x$，得到 $(\boldsymbol{E}-\boldsymbol{C})x=0$，可求得该齐次线性方程组的通解为 $x_1=0$，$x_2=0$，$x_3=k$，$x_4=0$，所以 $\boldsymbol{A}=0\boldsymbol{A}_1+0\boldsymbol{A}_2+k\boldsymbol{A}_3+0\boldsymbol{A}_4=k\boldsymbol{A}_3=k\begin{pmatrix}1&1\\1&0\end{pmatrix}$.（$k$ 任意）

【例 1.3.16】 $(x_1,\ x_2,\ x_3,\ x_4)^{\mathrm{T}}$ 是 $\boldsymbol{P}_3[t]$ 中的多项式 $f(t)$ 在基 $f_1(t)=4+4t+3t^2+t^3$，$f_2(t)=7+7t+5t^2+2t^3$，$f_3(t)=2-5t-3t^2-3t^3$，$f_4(t)=-3+8t+5t^2+5t^3$ 下的坐标；$(y_1,\ y_2,\ y_3,\ y_4)^{\mathrm{T}}$ 是 $f(t)$ 在基 $g_1(t),g_2(t),g_3(t),g_4(t)$ 下的坐标；且 $y_1=3x_1+5x_2$，$y_2=x_1+2x_2$，$y_3=2x_3-3x_4$，$y_4=-5x_3+8x_4$

试求：(1) 由基 $g_1(t),g_2(t),g_3(t),g_4(t)$ 到基 $f_1(t),f_2(t),f_3(t),f_4(t)$ 的过渡矩阵；

(2) 求基 $g_1(t),g_2(t),g_3(t),g_4(t)$；

(3) 求多项式 $g(t)=t^3-t^2+t-1$ 在基 $g_1(t),g_2(t),g_3(t),g_4(t)$ 下的坐标.

解 (1) 根据坐标之间的关系，写出坐标变换公式

$$\begin{pmatrix}y_1\\y_2\\y_3\\y_4\end{pmatrix}=\begin{pmatrix}3&5&0&0\\1&2&0&0\\0&0&2&-3\\0&0&-5&8\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}$$

从而得基变换公式为

$$(f_1,f_2,f_3,f_4)=(g_1,g_2,g_3,g_4)\boldsymbol{C}$$

其中 $\boldsymbol{C}=\begin{pmatrix}3&5&0&0\\1&2&0&0\\0&0&2&-3\\0&0&-5&8\end{pmatrix}$ 为基 g_1，g_2，g_3，g_4 到基 f_1，f_2，f_3，f_4 的过渡矩阵.

(2) 由 $(g_1,g_2,g_3,g_4)=(f_1,f_2,f_3,f_4)\boldsymbol{C}^{-1}=(f_1,f_2,f_3,f_4)\begin{pmatrix}2&-5&0&0\\-1&3&0&0\\0&0&8&3\\0&0&5&2\end{pmatrix}$，

得到

$$g_1(t)=2f_1(t)-f_2(t)=t^2+t+1,g_2(t)=-5f_1(t)+3f_2(t)=t^3+t+1,$$
$$g_3(t)=8f_3(t)+5f_4(t)=t^3+t^2+1,g_4(t)=3f_3(t)+2f_4(t)=t^3+t^2+t$$

(3) 设 $g(t)=x_1g_1(t)+x_2g_2(t)+x_3g_3(t)+x_4g_4(t)$，比较两边同次幂的系数得到

$$\begin{cases}x_2+x_3+x_4=1\\x_1+x_3+x_4=-1\\x_1+x_2+x_4=1\\x_1+x_2+x_3=-1\end{cases}$$

解得 $x_1=-1$，$x_2=1$，$x_3=-1$，$x_4=1$. 所以 $g(t)=t^3-t^2+t-1$ 在基 $g_1(t),g_2(t)$, $g_3(t),g_4(t)$ 下的坐标为 $(-1, 1, -1, 1)^{\mathrm{T}}$.

§1.4 线 性 子 空 间

前面我们讨论了线性空间的定义及其基、维数、坐标. 本节将对线性空间的子空间做一些介绍.

1. 线性子空间

定义 1.4.1 设 $\boldsymbol{V}$ 是数域 P 上的一个线性空间，S 为 $\boldsymbol{V}$ 中的一个非空子集. 若 S 对于 $\boldsymbol{V}$ 中所定义的加法与数乘法也构成一个线性空间，则 S 叫做 $\boldsymbol{V}$ 的一个线性子空间.

定理 1.4.1 线性空间 $\boldsymbol{V}$ 的非空子集 S 构成子空间的充分且必要条件是：S 对于 $\boldsymbol{V}$ 中的线性运算具有封闭性.

由定义 1.4.1 可知如下事实：如果 S 是 $\boldsymbol{V}$ 的一个线性子空间，则必有 $\dim S\leqslant\dim\boldsymbol{V}$. 单独一个零元素做成的子集 $\{\boldsymbol{0}\}$ 与子集 $\boldsymbol{V}$ 称为线性空间 $\boldsymbol{V}$ 的平凡子空间.

如果 $\boldsymbol{V}$ 是数域 P 上的一个线性空间，$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n\in\boldsymbol{V}$，我们考虑由 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 的一切线性组合所构成的集合 S，显然此集合非空.

又设 $\boldsymbol{\alpha}=a_1\boldsymbol{\alpha}_1+a_2\boldsymbol{\alpha}_2+\cdots+a_n\boldsymbol{\alpha}_n$，$\boldsymbol{\beta}=b_1\boldsymbol{\alpha}_1+b_2\boldsymbol{\alpha}_2+\cdots+b_n\boldsymbol{\alpha}_n\in S$.

则对于任意的 a，$b\in P$，$a\boldsymbol{\alpha}+b\boldsymbol{\beta}=\sum\limits_{i=1}^{n}(aa_i+bb_i)\boldsymbol{\alpha}_i\in S$.

$\therefore\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 的一切线性组合作成线性空间 V 的一个子空间 S. 我们称这个子空间为由 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 所生成的子空间，记作 $S=\mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$，其中 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 称为子

空间 S 的一组生成元.

注 即使 S 不是 $\boldsymbol{V}$ 的子空间，$\mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ $(\forall \boldsymbol{\alpha}_i \in S)$ 总是作成 $\boldsymbol{V}$ 的子空间.

【例 1.4.1】 考虑 $\mathbf{R}^n$ 中的标准基 $\boldsymbol{\varepsilon}_i=(0, \cdots, 0, \underset{(i)}{1}, 0, \cdots, 0)^{\mathrm{T}}$ $(i=1, 2, \cdots, n)$，对任意的 $\boldsymbol{\alpha}=(a_1, a_2, \cdots, a_n) \in \mathbf{R}^n$，有 $\boldsymbol{\alpha}=a_1\boldsymbol{\varepsilon}_1+a_2\boldsymbol{\varepsilon}_2+\cdots+a_n\boldsymbol{\varepsilon}_n$，所以 $\mathbf{R}^n=\mathrm{Span}\{\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n\}$，由此可知 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n$ 是 $\mathbf{R}^n$ 的一组生成元. 显然，$\mathbf{R}^n$ 的任何一组基均可生成 $\mathbf{R}^n$.

【例 1.4.2】 在 $\boldsymbol{P}_n[x]$ 中所有次数不大于 $r-1(r\leqslant n)$ 的多项式的全体 $\boldsymbol{P}_{r-1}[x]$ 构成 $\boldsymbol{P}_n[x]$ 的一个 r 维子空间，并且 $\boldsymbol{P}_{r-1}[x]=\mathrm{Span}\{1,x,x^2,\cdots,x^{r-1}\}$.

【例 1.4.3】 在 $\boldsymbol{R}^n$ 中，前 k 个分量为 0 的一切 n 元数组 $(0, \cdots 0, x_{k+1}, x_{k+2}, \cdots, x_n)^{\mathrm{T}}$ $(k\leqslant n)$ 的全体构成一个 $n-k$ 维子空间，且此子空间即为 $\mathrm{Span}\{\boldsymbol{\varepsilon}_{k+1}, \boldsymbol{\varepsilon}_{k+2}, \cdots, \boldsymbol{\varepsilon}_n\}$.

【例 1.4.4】 齐次线性方程组 $\boldsymbol{Ax}=0$ 的全体解向量构成一个子空间，我们称这个子空间为该方程组的解空间，其基是该方程组的基解，维数为 $n-r$，其中 r 为系数矩阵的秩.

【例 1.4.5】 $\mathbf{R}^{2\times2}$ 的子集 $\boldsymbol{W}_1=\{\boldsymbol{A} \mid |\boldsymbol{A}|=0, \boldsymbol{A}\in \mathbf{R}^{2\times2}\}$ 是否构成子空间?为什么?

解 不构成. 取 $\boldsymbol{A}=\begin{pmatrix}1&0\\0&0\end{pmatrix}$，$\boldsymbol{B}=\begin{pmatrix}0&0\\0&1\end{pmatrix}$，则 $|\boldsymbol{A}|=|\boldsymbol{B}|=0$，即 $\boldsymbol{A}, \boldsymbol{B}\in\boldsymbol{W}_1$.

但是 $\boldsymbol{A}+\boldsymbol{B}=\begin{pmatrix}1&0\\0&1\end{pmatrix}$，$|\boldsymbol{A}+\boldsymbol{B}|=1$. 可见 $\boldsymbol{A}+\boldsymbol{B}\notin\boldsymbol{W}_1$，故 $\boldsymbol{W}_1$ 不构成子空间.

2. 子空间的交与和

定理 1.4.2 设 $\boldsymbol{V}_1$，$\boldsymbol{V}_2$ 是线性空间 $\boldsymbol{V}$ 的子空间，则 $\boldsymbol{V}_1\cap\boldsymbol{V}_2$ 也是 $\boldsymbol{V}$ 的子空间，称这个子空间为 $\boldsymbol{V}_1$ 与 $\boldsymbol{V}_2$ 的交.

证 由于 V_1，V_2 是线性空间 V 的子空间，$\mathbf{0}\in\boldsymbol{V}_1$，$\mathbf{0}\in\boldsymbol{V}_2$，因此 $\mathbf{0}\in\boldsymbol{V}_1\cap\boldsymbol{V}_2$，

$\therefore \boldsymbol{V}_1\cap\boldsymbol{V}_2\neq\boldsymbol{\Phi}$.

对任意的 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\boldsymbol{V}_1\cap\boldsymbol{V}_2$，则 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\boldsymbol{V}_1$ 并且 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\boldsymbol{V}_2$，$\therefore \boldsymbol{\alpha}+\boldsymbol{\beta}\in\boldsymbol{V}_1$，$a\boldsymbol{\alpha}\in\boldsymbol{V}_1$；

$\boldsymbol{\alpha}+\boldsymbol{\beta}\in\boldsymbol{V}_2$，$a\boldsymbol{\alpha}\in\boldsymbol{V}_2$. $\therefore \boldsymbol{\alpha}+\boldsymbol{\beta}\in\boldsymbol{V}_1\cap\boldsymbol{V}_2$，$a\boldsymbol{\alpha}\in\boldsymbol{V}_1\cap\boldsymbol{V}_2$，所以 $\boldsymbol{V}_1\cap\boldsymbol{V}_2$ 是 $\boldsymbol{V}$ 的子空间.

【例 1.4.6】 设 $\boldsymbol{V}_1$，$\boldsymbol{V}_2$ 分别为齐次线性方程组 $\boldsymbol{Ax}=0$，$\boldsymbol{Bx}=0$ 的解空间，则 $\boldsymbol{V}_1\cap\boldsymbol{V}_2$ 为 $\begin{cases}\boldsymbol{Ax}=0\\\boldsymbol{Bx}=0\end{cases}$ 的解空间.

定理 1.4.3 设 $\boldsymbol{V}_1$，$\boldsymbol{V}_2$ 是线性空间 $\boldsymbol{V}$ 的子空间，则 $\boldsymbol{V}_1+\boldsymbol{V}_2=\{\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2 \mid \forall\boldsymbol{\alpha}_1\in\boldsymbol{V}_1, \boldsymbol{\alpha}_2\in\boldsymbol{V}_2\}$ 也是 $\boldsymbol{V}$ 的子空间，称这个子空间为 $\boldsymbol{V}_1$ 与 $\boldsymbol{V}_2$ 的和.

证 显然 $\boldsymbol{V}_1+\boldsymbol{V}_2$ 非空，对任意的 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\boldsymbol{V}_1+\boldsymbol{V}_2$，有

$$\boldsymbol{\alpha}=\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2, \boldsymbol{\beta}=\boldsymbol{\beta}_1+\boldsymbol{\beta}_2 \quad (\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i\in\boldsymbol{V}_i, i=1,2)$$

$$\therefore \boldsymbol{\alpha}+\boldsymbol{\beta}=(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2)+(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)=(\boldsymbol{\alpha}_1+\boldsymbol{\beta}_1)+(\boldsymbol{\alpha}_2+\boldsymbol{\beta}_2)\in\boldsymbol{V}_1+\boldsymbol{V}_2.$$

又对任意的 $a\in P$，$a\boldsymbol{\alpha}=a\boldsymbol{\alpha}_1+a\boldsymbol{\alpha}_2\in\boldsymbol{V}_1+\boldsymbol{V}_2$，所以 $\boldsymbol{V}_1+\boldsymbol{V}_2$ 是 $\boldsymbol{V}$ 的子空间.

如果线性空间 $\boldsymbol{V}_1$ 与 $\boldsymbol{V}_2$ 分别由 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s$ 与 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t$ 所生成，那么

$$\begin{aligned}\boldsymbol{V}_1+\boldsymbol{V}_2&=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_s\}+\mathrm{Span}\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_t\}\\&=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_s,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_t\}\end{aligned}$$

$$\boldsymbol{V}_1\cap\boldsymbol{V}_2=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_s\}\cap\mathrm{Span}\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_t\}=\mathrm{Span}\{\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_r\}$$

式中，$\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_r\in\boldsymbol{V}_1\cap\boldsymbol{V}_2$，$r\leqslant s, t$.

以上运算可以推广到 n 个子空间 $\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots, \boldsymbol{V}_n$.

【例 1.4.7】 设 $\boldsymbol{\alpha}_1=(1,2,1,0)$，$\boldsymbol{\alpha}_2=(-1,1,1,1)$，$\boldsymbol{\beta}_1=(2,-1,0,1)$，$\boldsymbol{\beta}_2=(1,-1,3,7)$，求 $V_1=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\}$ 与 $V_2=\mathrm{Span}\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\}$ 的和与交的维数以及它们的基.

解 $\because \boldsymbol{V}_1+\boldsymbol{V}_2=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\}+\mathrm{Span}\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\}=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\}$

向量组 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$ 的秩为 3，并且 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\boldsymbol{\beta}_1$ 为其一个极大线性无关组.

$\therefore \dim(\boldsymbol{V}_1+\boldsymbol{V}_2)=3$，并且 $\boldsymbol{V}_1+\boldsymbol{V}_2$ 的一个基为 $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1\}$，即

$$\boldsymbol{V}_1+\boldsymbol{V}_2=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1\}$$

对 $\boldsymbol{V}_1\cap\boldsymbol{V}_2$ 而言，对于任意的 $\boldsymbol{\alpha}\in\boldsymbol{V}_1\cap\boldsymbol{V}_2$，有 a_1，a_2，b_1，$b_2\in P$，使 $\boldsymbol{\alpha}=a_1\boldsymbol{\alpha}_1+a_2\boldsymbol{\alpha}_2=b_1\boldsymbol{\beta}_1+b_2\boldsymbol{\beta}_2$，所以 $a_1\boldsymbol{\alpha}_1+a_2\boldsymbol{\alpha}_2-b_1\boldsymbol{\beta}_1-b_2\boldsymbol{\beta}_2=0$.

则 $(a_1,2a_1,a_1,0)+(-a_2,a_2,a_2,a_2)-(2b_1,-b_1,0,b_1)-(b_2,-b_2,3b_2,7b_2)=(0,0,0,0)$.

$$\therefore\begin{cases}a_1-a_2-2b_1-b_2=0\\2a_1+a_2+b_1+b_2=0\\a_1+a_2\qquad\quad-3b_2=0\\a_2-b_1-7b_2=0\end{cases}$$，其基础解为$(1,-4,3,-1)^{\mathrm{T}}$.

又 $\because \boldsymbol{\alpha}=\boldsymbol{\alpha}_1-4\boldsymbol{\alpha}_2=3\boldsymbol{\beta}_1-\boldsymbol{\beta}_2=(5,-2,-3,-4)$

$\therefore \dim(\boldsymbol{V}_1\cap\boldsymbol{V}_2)=1$，则

$\boldsymbol{\alpha}=(5,-2,-3,-4)$ 是 $\boldsymbol{V}_1\cap\boldsymbol{V}_2$ 的一个基，即 $\boldsymbol{V}_1\cap\boldsymbol{V}_2=\mathrm{Span}\{\boldsymbol{\alpha}\}$.

以上给出了子空间的定义，并讨论了子空间的交与和，为了对子空间有进一步的了解，我们将深入讨论子空间的基和维数.

3. 子空间的基与维数

【例 1.4.8】 求 $\mathbf{R}^{2\times3}$ 的子空间 $\boldsymbol{W}=\left\{\begin{pmatrix}a&b&0\\c&0&d\end{pmatrix}\middle|a+b+d=0,\ a,b,c,d\in\mathbf{R}\right\}$ 的基与维数.

解 $\boldsymbol{W}$ 中任一矩阵 $\boldsymbol{A}=\begin{pmatrix}a&b&0\\c&0&d\end{pmatrix}$ 的元素 a，b，c，d 满足齐次方程组 $a+b+0c+d=0$

该方程组的通解为 $a=-k_1-k_3$，$b=k_1$，$c=k_2$，$d=k_3$，$(k_1,k_2,k_3\in\mathbf{R})$

从而 $\boldsymbol{A}=\begin{bmatrix}-k_1-k_3&k_1&0\\k_2&0&k_3\end{bmatrix}=k_1\boldsymbol{A}_1+k_2\boldsymbol{A}_2+k_3\boldsymbol{A}_3$，其中 $\boldsymbol{A}_1=\begin{pmatrix}-1&1&0\\0&0&0\end{pmatrix}$，$\boldsymbol{A}_2=\begin{pmatrix}0&0&0\\1&0&0\end{pmatrix}$，$\boldsymbol{A}_3=\begin{pmatrix}-1&0&0\\0&0&1\end{pmatrix}$.

即 $\boldsymbol{A}$ 可由 $\boldsymbol{A}_1$，$\boldsymbol{A}_2$，$\boldsymbol{A}_3$ 线性表示，易证 $\boldsymbol{A}_1$，$\boldsymbol{A}_2$，$\boldsymbol{A}_3$ 线性无关，所以 $\boldsymbol{W}$ 是 3 维的，并且 $\boldsymbol{A}_1$，$\boldsymbol{A}_2$，$\boldsymbol{A}_3$ 是 $\boldsymbol{W}$ 的一组基.

【例 1.4.9】 已知 $\boldsymbol{A}=\begin{pmatrix}1&3\\0&2\end{pmatrix}$，求 $\mathbf{R}^{2\times2}$ 的子空间 $\boldsymbol{W}=\{\boldsymbol{X}\mid\boldsymbol{AX}=\boldsymbol{XA},\boldsymbol{X}\in\mathbf{R}^{2\times2}\}$ 的基与维数.

解 设 $\boldsymbol{X}=\begin{pmatrix}x_1&x_2\\x_3&x_4\end{pmatrix}\in\boldsymbol{W}$，则由 $\boldsymbol{AX}=\boldsymbol{XA}$，得到

$$\begin{pmatrix} x_1+3x_3 & x_2+3x_4 \\ 2x_3 & 2x_4 \end{pmatrix} = \begin{pmatrix} x_1 & 3x_1+2x_2 \\ x_3 & 3x_3+2x_4 \end{pmatrix}$$

比较分量得齐次方程组

$$3x_3=0, -3x_1-x_2+3x_4=0, x_3=0, -3x_3=0$$

此方程组得通解为

$$x_1=-k_1+k_2, x_2=3k_1, x_3=0, x_4=k_2; (k_1, k_2 \in \mathbf{R})$$

于是

$$\boldsymbol{X}=\begin{pmatrix} -k_1+k_2 & 3k_1 \\ 0 & k_2 \end{pmatrix}=k_1\begin{pmatrix} -1 & 3 \\ 0 & 0 \end{pmatrix}+k_2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

令 $\boldsymbol{A}_1=\begin{pmatrix} -1 & 3 \\ 0 & 0 \end{pmatrix}$，$\boldsymbol{A}_2=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$，则 $\boldsymbol{A}_1$，$\boldsymbol{A}_2 \in \boldsymbol{W}$，易证 $\boldsymbol{A}_1$，$\boldsymbol{A}_2$ 线性无关，故 $\boldsymbol{W}$ 是 2 维的，并且 $\boldsymbol{A}_1$，$\boldsymbol{A}_2$ 是 $\boldsymbol{W}$ 的一组基.

定义 1.4.2 设$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$与$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$是线性空间 $\mathbf{V}$ 中的两个向量组，如果对于任意的 $\boldsymbol{\alpha}_i (i=1,2,\cdots,s)$ 均可以由向量组$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$线性表示，而任意的 $\boldsymbol{\beta}_j (j=1,2,\cdots,t)$ 均可以由向量组$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$线性表示，则称这两个向量组等价.

定理 1.4.4 设向量组$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$线性无关，并且每一个 $\boldsymbol{\alpha}_i$ 都可以由向量组$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$线性表示，那么必有 $s \leqslant t$，并且在必要时可以对$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$中的向量重新编号，使得用 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s$ 替换 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t$ 后，所得的向量组 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s, \boldsymbol{\beta}_{s+1}, \boldsymbol{\beta}_{s+2}, \cdots, \boldsymbol{\beta}_t\}$ 与$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$等价.

证 对$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$中的向量的个数 s 应用数学归纳法.

当 $s=0$ 时，论断显然正确.

假设 $s>0$，并且对向量组$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$中含有 $s-1$ 个向量的情形，结论成立.

当向量组$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$中含有 s 个向量时：

$\because \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s$ 线性无关.

$\therefore \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{s-1}$ 也线性无关.

又归纳假设 $s-1 \leqslant t$，并且以 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{s-1}$ 替换$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$中的前 $s-1$ 个向量，得到一个与$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$等价的向量组 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{s-1}, \boldsymbol{\beta}_s, \boldsymbol{\beta}_{s+1}, \cdots, \boldsymbol{\beta}_t\}$，由于 $\boldsymbol{\alpha}_s$ 可以由$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$线性表示，所以它也可以由与$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$等价的向量组$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{s-1}, \boldsymbol{\beta}_s, \boldsymbol{\beta}_{s+1}, \cdots, \boldsymbol{\beta}_t\}$ 线性表示.

$\therefore \boldsymbol{\alpha}_s = a_1\boldsymbol{\alpha}_1 + a_2\boldsymbol{\alpha}_2 + \cdots a_{s-1}\boldsymbol{\alpha}_{s-1} + b_s\boldsymbol{\beta}_s + b_{s+1}\boldsymbol{\beta}_{s+1} + \cdots + b_t\boldsymbol{\beta}_t$.

如果所有的 b_j 都为零，则上式变为 $\boldsymbol{\alpha}_s = a_1\boldsymbol{\alpha}_1 + a_2\boldsymbol{\alpha}_2 + \cdots a_{s-1}\boldsymbol{\alpha}_{s-1}$，即 $\boldsymbol{\alpha}_s$ 可以由 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{s-1}$ 线性表示，而这与$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$线性无关的前提矛盾.

因此，至少有一个 $b_j \neq 0$，$\therefore s-1<t$，$\therefore s \leqslant t$.

适当地对 $\boldsymbol{\beta}_s, \boldsymbol{\beta}_{s+1}, \cdots, \boldsymbol{\beta}_t$ 编号，不妨设 $b_s \neq 0$，则有

$$\boldsymbol{\beta}_s = \sum_{i=1}^{s-1} -\left(\frac{a_i}{b_s}\right)\boldsymbol{\alpha}_i + \frac{\boldsymbol{\alpha}_s}{b_s} + \sum_{j=s+1}^{t} -\left(\frac{b_j}{b_s}\right)\boldsymbol{\beta}_j$$

即 $\boldsymbol{\beta}_s$ 可以由 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s, \boldsymbol{\beta}_{s+1}, \boldsymbol{\beta}_{s+2}, \cdots, \boldsymbol{\beta}_t\}$ 线性表示.

同时，$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{s-1}, \boldsymbol{\beta}_s, \boldsymbol{\beta}_{s+1}, \cdots, \boldsymbol{\beta}_t\}$ 中的元素除去 $\boldsymbol{\beta}_s$ 外，其余所有向量都在 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s, \boldsymbol{\beta}_{s+1}, \boldsymbol{\beta}_{s+2}, \cdots, \boldsymbol{\beta}_t\}$ 中出现，所以 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{s-1}, \boldsymbol{\beta}_s, \boldsymbol{\beta}_{s+1}, \cdots, \boldsymbol{\beta}_t\}$

中所有向量都可以由 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s, \boldsymbol{\beta}_{s+1}, \boldsymbol{\beta}_{s+2}, \cdots, \boldsymbol{\beta}_t\}$ 线性表示.

同理，$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s, \boldsymbol{\beta}_{s+1}, \boldsymbol{\beta}_{s+2}, \cdots, \boldsymbol{\beta}_t\}$ 中所有向量都可以由 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{s-1}, \boldsymbol{\beta}_s, \boldsymbol{\beta}_{s+1}, \cdots, \boldsymbol{\beta}_t\}$ 线性表示，因此这两个向量组等价.

由归纳假设，这两个向量组与 $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$ 等价.

推论 1.4.1 两个等价的线性无关向量组含有相同个数的向量.

证 设 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$ 与 $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$ 是两个等价的线性无关向量组，由定理 1.4.4，有 $s \leqslant t$，且 $t \leqslant s$，则 $s=t$

定理 1.4.5 两个向量组 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$ 与 $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$ 生成的子空间相同的充分且必要条件为 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$ 与 $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$ 等价.

证 必要性：若 $\mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\} = \mathrm{Span}\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$，

则 $\boldsymbol{\alpha}_i \in \mathrm{Span}\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$，并且 $\boldsymbol{\beta}_j \in \mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$，即向量组 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$ 与 $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$ 可以互相表示，所以这两个向量组等价.

充分性：若向量组 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$ 与 $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$ 等价，则

$$\forall \boldsymbol{\alpha} \in \mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$$

均有 $\boldsymbol{\alpha} = \sum_{i=1}^{s} a_i \boldsymbol{\alpha}_i = \sum_{j=1}^{t} b_j \boldsymbol{\beta}_j \in \mathrm{Span}\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$；

同理 $\forall \boldsymbol{\beta} \in \mathrm{Span}\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$，均有 $\boldsymbol{\beta} = \sum_{j=1}^{t} b_j \boldsymbol{\beta}_j = \sum_{i=1}^{s} a_i \boldsymbol{\alpha}_i \in \mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$，

$\therefore \mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\} = \mathrm{Span}\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_t\}$

定理 1.4.6 线性空间 $\boldsymbol{V} = \mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$ 的维数 $\dim\boldsymbol{V}$ 等于向量组 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$ 的秩.

证 设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s$ 的秩为 r，并设 $\boldsymbol{\alpha}_{i_1}, \boldsymbol{\alpha}_{i_2}, \cdots, \boldsymbol{\alpha}_{i_r}$ 是它的一个极大线性无关向量组，则 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s$ 与 $\boldsymbol{\alpha}_{i_1}, \boldsymbol{\alpha}_{i_2}, \cdots, \boldsymbol{\alpha}_{i_r}$ 等价，所以 $\forall \boldsymbol{\alpha} \in \mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$，均有 $\boldsymbol{\alpha} = a_{i_1}\boldsymbol{\alpha}_{i_1} + a_{i_2}\boldsymbol{\alpha}_{i_2} + \cdots + a_{i_r}\boldsymbol{\alpha}_{i_r}$，则 $\dim\boldsymbol{V} = r$.

定理 1.4.7 设 $\boldsymbol{V}^r$ 是 n 维线性空间 $\boldsymbol{V}^n$ 的一个 r 维子空间，$\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r$ 是 $\boldsymbol{V}^r$ 的一个基，则 $\boldsymbol{V}^n$ 中存在 $n-r$ 个向量 $\boldsymbol{\alpha}_{r+1}, \boldsymbol{\alpha}_{r+2}, \cdots, \boldsymbol{\alpha}_n$，使得 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r, \boldsymbol{\alpha}_{r+1}, \boldsymbol{\alpha}_{r+2}, \cdots, \boldsymbol{\alpha}_n$ 为 $\boldsymbol{V}^n$ 的一个基. 特别是，n 维线性空间 $\boldsymbol{V}^n$ 中任意 n 个线性无关的向量都可以取做基.

证 设 $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n\}$ 是 $\boldsymbol{V}^n$ 的一个基，则每一个 $\boldsymbol{\alpha}_i$ 都可以由 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n$ 线性表示，又因为 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r$ 线性无关，所以由定理 1.4.4，适当地对 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n$ 编号，就可以用 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r$ 替换前 r 个基向量 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_r$，得到一个与 $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n\}$ 等价的向量组 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r, \boldsymbol{\beta}_{r+1}, \boldsymbol{\beta}_{r+2}, \cdots, \boldsymbol{\beta}_n\}$，这个向量组也含有 n 个线性无关的向量，它也构成 $\boldsymbol{V}^n$ 的一个基，取 $\boldsymbol{\alpha}_j = \boldsymbol{\beta}_j\,(j=r+1, r+2, \cdots, n)$ 即可.

定理 1.4.8 设 $\boldsymbol{V}_1$，$\boldsymbol{V}_2$ 是线性空间 $\boldsymbol{V}$ 的任意两个子空间，则

$$\dim(\boldsymbol{V}_1 + \boldsymbol{V}_2) = \dim\boldsymbol{V}_1 + \dim\boldsymbol{V}_2 - \dim(\boldsymbol{V}_1 \cap \boldsymbol{V}_2)$$

证 设 $\dim\boldsymbol{V}_1 = n_1$，$\dim\boldsymbol{V}_2 = n_2$，$\dim(\boldsymbol{V}_1 \cap \boldsymbol{V}_2) = r$.

取 $\boldsymbol{V}_1 \cap \boldsymbol{V}_2$ 的一个基为 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r$；由定理 1.4.7，将其扩充为 $\boldsymbol{V}_1$ 与 $\boldsymbol{V}_2$ 的一个基，则

$$\boldsymbol{V}_1 = \mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_{n_1-r}\}$$
$$\boldsymbol{V}_2 = \mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_{n_2-r}\}$$

$\therefore \boldsymbol{V}_1+\boldsymbol{V}_2=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_{n_1-r},\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_{n_2-r}\}$.

以下证明这组生成元线性无关：

令 $h_1\boldsymbol{\alpha}_1+h_2\boldsymbol{\alpha}_2+\cdots+h_r\boldsymbol{\alpha}_r+k_1\boldsymbol{\beta}_1+k_2\boldsymbol{\beta}_2+\cdots+k_{n_1-r}\boldsymbol{\beta}_{n_1-r}+l_1\boldsymbol{\gamma}_1+l_2\boldsymbol{\gamma}_2+\cdots+l_{n_2-r}\boldsymbol{\gamma}_{n_2-r}=\boldsymbol{0}$

再令

$$\boldsymbol{\alpha}=h_1\boldsymbol{\alpha}_1+h_2\boldsymbol{\alpha}_2+\cdots+h_r\boldsymbol{\alpha}_r+k_1\boldsymbol{\beta}_1+k_2\boldsymbol{\beta}_2+\cdots+k_{n_1-r}\boldsymbol{\beta}_{n_1-r},$$

则 $\boldsymbol{\alpha}=-l_1\boldsymbol{\gamma}_1-l_2\boldsymbol{\gamma}_2-\cdots-l_{n_2-r}\boldsymbol{\gamma}_{n_2-r}$.

$\boldsymbol{\alpha}\in\boldsymbol{V}_1$ 且 $\boldsymbol{\alpha}\in\boldsymbol{V}_2$，$\therefore\boldsymbol{\alpha}\in\boldsymbol{V}_1\cap\boldsymbol{V}_2$，则 $\boldsymbol{\alpha}=p_1\boldsymbol{\alpha}_1+p_2\boldsymbol{\alpha}_2+\cdots+p_r\boldsymbol{\alpha}_r$，

$\therefore\boldsymbol{\alpha}=p_1\boldsymbol{\alpha}_1+p_2\boldsymbol{\alpha}_2+\cdots+p_r\boldsymbol{\alpha}_r=-l_1\boldsymbol{\gamma}_1-l_2\boldsymbol{\gamma}_2-\cdots-l_{n_2-r}\boldsymbol{\gamma}_{n_2-r}$

即 $p_1\boldsymbol{\alpha}_1+p_2\boldsymbol{\alpha}_2+\cdots+p_r\boldsymbol{\alpha}_r+l_1\boldsymbol{\gamma}_1+l_2\boldsymbol{\gamma}_2+\cdots+l_{n_2-r}\boldsymbol{\gamma}_{n_2-r}=\boldsymbol{0}$.

又$\because\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\cdots$，$\boldsymbol{\alpha}_r$，$\boldsymbol{\gamma}_1$，$\boldsymbol{\gamma}_2$，$\cdots$，$\boldsymbol{\gamma}_{n_2-r}$线性无关，

$\therefore$ $p_1=p_2=\cdots=p_r=l_1=l_2=\cdots=l_{n_2-r}=0$，$\boldsymbol{\alpha}=\boldsymbol{0}$，

则 $h_1\boldsymbol{\alpha}_1+h_2\boldsymbol{\alpha}_2+\cdots+h_r\boldsymbol{\alpha}_r+k_1\boldsymbol{\beta}_1+k_2\boldsymbol{\beta}_2+\cdots+k_{n_1-r}\boldsymbol{\beta}_{n_1-r}=\boldsymbol{0}$.

又$\because\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\cdots$，$\boldsymbol{\alpha}_r$，$\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，$\cdots$，$\boldsymbol{\beta}_{n_1-r}$也线性无关.

$\therefore$ $h_1=h_2=\cdots=h_r=k_1=k_2=\cdots=k_{n_1-r}=0$.

则$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\cdots$，$\boldsymbol{\alpha}_r$，$\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，$\cdots$，$\boldsymbol{\beta}_{n_1-r}$，$\boldsymbol{\gamma}_1$，$\boldsymbol{\gamma}_2$，$\cdots$，$\boldsymbol{\gamma}_{n_2-r}$线性无关，作成 $\boldsymbol{V}_1+\boldsymbol{V}_2$ 的一个基. 因此，$\dim(\boldsymbol{V}_1+\boldsymbol{V}_2)=n_1+n_2-r=\dim\boldsymbol{V}_1+\dim\boldsymbol{V}_2-\dim(\boldsymbol{V}_1\cap\boldsymbol{V}_2)$.

推论 1.4.2 若 $\boldsymbol{V}_1$，$\boldsymbol{V}_2$ 是线性空间 $\boldsymbol{V}^n$ 的两个子空间，$\dim\boldsymbol{V}^n=n$，并且 $\dim\boldsymbol{V}_1+\dim\boldsymbol{V}_2>n$，则必存在 $\boldsymbol{\alpha}\neq\boldsymbol{0}$，且 $\boldsymbol{\alpha}\in\boldsymbol{V}_1\cap\boldsymbol{V}_2$.

证 因为 $\boldsymbol{V}_1$，$\boldsymbol{V}_2$ 是 $\boldsymbol{V}^n$ 的子空间，所以 $\dim(\boldsymbol{V}_1+\boldsymbol{V}_2)=\dim\boldsymbol{V}_1+\dim\boldsymbol{V}_2-\dim(\boldsymbol{V}_1\cap\boldsymbol{V}_2)$，则

$$\dim(\boldsymbol{V}_1\cap\boldsymbol{V}_2)=\dim\boldsymbol{V}_1+\dim\boldsymbol{V}_2-\dim(\boldsymbol{V}_1+\boldsymbol{V}_2)$$

又$\because\dim\boldsymbol{V}_1+\dim\boldsymbol{V}_2>n$，

$\therefore\dim\boldsymbol{V}_1+\dim\boldsymbol{V}_2-\dim(\boldsymbol{V}_1+\boldsymbol{V}_2)>n-\dim(\boldsymbol{V}_1+\boldsymbol{V}_2)$，

$\because(\boldsymbol{V}_1+\boldsymbol{V}_2)$是 V^n 的子空间，所以 $\dim(\boldsymbol{V}_1+\boldsymbol{V}_2)\leqslant\dim V^n=n$，

$\therefore\dim(\boldsymbol{V}_1\cap\boldsymbol{V}_2)=\dim\boldsymbol{V}_1+\dim\boldsymbol{V}_2-\dim(\boldsymbol{V}_1+\boldsymbol{V}_2)>n-\dim(\boldsymbol{V}_1+\boldsymbol{V}_2)\geqslant0$

则 $\dim(\boldsymbol{V}_1\cap\boldsymbol{V}_2)>0$，所以 $\boldsymbol{V}_1\cap\boldsymbol{V}_2$ 为非零子空间，必含有非零向量.

【例 1.4.10】 设 $\mathbf{R}^{2\times2}$的两个子空间为

$$\boldsymbol{W}_1=\left\{\boldsymbol{A}=\begin{pmatrix}x_1&x_2\\x_3&x_4\end{pmatrix}\middle|\,x_1-x_2+x_3-x_4=0\right\}$$

$$\boldsymbol{W}_2=\mathrm{span}\{\boldsymbol{B}_1,\boldsymbol{B}_2\},\boldsymbol{B}_1=\begin{pmatrix}1&0\\2&3\end{pmatrix},\boldsymbol{B}_2=\begin{pmatrix}1&-1\\0&1\end{pmatrix}$$

求 $\boldsymbol{W}_1+\boldsymbol{W}_2$，$\boldsymbol{W}_1\cap\boldsymbol{W}_2$.

解 齐次线性方程组 $x_1-x_2+x_3-x_4=0$ 的通解为

$$x_1=t_1-t_2+t_3,x_2=t_1,x_3=t_2,x_4=t_3\quad(t_1,t_2,t_3\in\mathbf{R})$$

于是 $\boldsymbol{A}\in\boldsymbol{W}_1$ 可表示为

$$\boldsymbol{A}=\begin{pmatrix}t_1-t_2+t_3&t_1\\t_2&t_3\end{pmatrix}=t_1\boldsymbol{A}_1+t_2\boldsymbol{A}_2+t_3\boldsymbol{A}_3$$

其中
$$\boldsymbol{A}_1=\begin{pmatrix}1&1\\0&0\end{pmatrix},\boldsymbol{A}_2=\begin{pmatrix}-1&0\\1&0\end{pmatrix},\boldsymbol{A}_3=\begin{pmatrix}1&0\\0&1\end{pmatrix}$$

从而 $W_1 = \mathrm{span}(A_1, A_2, A_3)$，且 $W_1 + W_2 = \mathrm{span}(A_1, A_2, A_3, B_1, B_2)$.

取 $\mathbf{R}^{2\times 2}$ 的基 $E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$，$E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$，$E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$，$E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$，以 A_1，A_2，A_3，B_1，B_2 在该基下的坐标为列向量构造矩阵 A，并对 A 作初等行变换

$$A = \begin{pmatrix} 1 & -1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 3 & 1 \end{pmatrix} \xrightarrow[r_1 - r_4]{\substack{r_1 - r_2 \\ r_1 + r_3}} \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 3 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

可见 $\mathrm{rank}A=4$，且 A 的第 1，2，3，5 列是 A 的列向量组的一个极大无关组. 从而 W_1+W_2 的维数为 4，且 A_1，A_2，A_3，B_2 为它的一组基.

设 $A \in W_1 \cap W_2$，则存在数组 k_1，k_2，k_3 和 l_1，l_2，使得

$$A = k_1 A_1 + k_2 A_2 + k_3 A_3 = l_1 B_1 + l_2 B_2$$

即

$$k_1 A_1 + k_2 A_2 + k_3 A_3 - l_2 B_1 - l_2 B_2 = 0$$

比较等式两边的对应元素得

$$\begin{cases} k_1 - k_2 + k_3 - l_1 - l_2 = 0 \\ k_1 + l_2 = 0 \\ k_2 - 2l_1 = 0 \\ k_3 - 3l_1 - l_2 = 0 \end{cases}$$

求得其通解为

$$k_1 = 0, k_2 = 2t, k_3 = 3t, l_1 = t, l_2 = 0 \quad (t \in \mathbf{R})$$

于是

$$A = l_1 B_1 + l_2 B_2 = tB_1 \quad (t \in \mathbf{R})$$

故 $W_1 \cap W_2$ 是 1 维的，且 B_1 是它的基.

定义 1.4.3 设 V_1，V_2 是线性空间 V 的两个子空间，若其和 V_1+V_2 中每个向量的分解式 $\alpha = \alpha_1 + \alpha_2 (\alpha_1 \in V_1, \alpha_2 \in V_2)$ 是唯一的，则称 V_1+V_2 为直和，记作 $V_1 \oplus V_2$.

定理 1.4.9 V_1+V_2 是直和的充分与必要条件是 $V_1 \cap V_2 = \{\mathbf{0}\}$.

证 必要性：设 V_1+V_2 是直和，$\because \mathbf{0} \in \{\mathbf{0}\}$ 并且 $0 \in V_1 \cap V_2$，$\therefore \{\mathbf{0}\} \subseteq V_1 \cap V_2$.

又$\because \forall \alpha \in V_1 \cap V_2$，$\alpha \in V_1$ 并且 $\alpha \in V_2$，而 $\mathbf{0} = \alpha + (-\alpha) \in V_1 \oplus V_2$，由定义 1.4.3 分解式的唯一性有 $\alpha = -\alpha = \mathbf{0}$，$\therefore V_1 \cap V_2 \subseteq \{\mathbf{0}\}$，$\therefore V_1 \cap V_2 = \{\mathbf{0}\}$.

充分性：设 $V_1 \cap V_2 = \{\mathbf{0}\}$，$\forall \alpha \in V_1 + V_2$，若有 $\alpha = \alpha_1 + \alpha_2 = \beta_1 + \beta_2$ (α_1，$\beta_1 \in V_1$，α_2，$\beta_2 \in V_2$)，则必有 $(\alpha_1 - \beta_1) = -(\alpha_2 - \beta_2)$，即 $(\alpha_1 - \beta_1) + (\alpha_2 - \beta_2) = \mathbf{0}$.

若 $\alpha_2 \neq \beta_2$，则 $(\alpha_1 - \beta_1) = (\beta_2 - \alpha_2) \neq \mathbf{0}$，并且 $\alpha_1 - \beta_1$，$\beta_2 - \alpha_2 \in V_1 \cap V_2$，这与假设矛盾.

$\therefore \alpha_1 = \beta_1$，$\alpha_2 = \beta_2$，由此可知 α 的分解式唯一，所以 V_1+V_2 是直和.

定理 1.4.10 V_1+V_2 是直和的充分与必要条件是 $\dim(V_1+V_2) = \dim V_1 + \dim V_2$.

证 由定理 1.4.8、定理 1.4.9 得证.

由此可知在 V_1，V_2 中分别取一个基 α_1，α_2，…，α_{r_1} 与 β_1，β_2，…，β_{r_2} 组成的向量组 $\{\alpha_1, \alpha_2, \cdots, \alpha_{r_1}, \beta_1, \beta_2, \cdots, \beta_{r_2}\}$ 即为 $V_1 \oplus V_2$ 的一个基.

定理 1.4.11 设 V_1 是线性空间 V^n 的一个子空间，则一定存在着 V^n 的另一个子空间 V_2，使得 $V^n = V_1 \oplus V_2$.

证 取 V_1 的一个基 α_1，α_2，…，α_r，将其扩充为 V^n 的一个基 $\{\alpha_1, \cdots, \alpha_r, \alpha_{r+1}, \cdots,$

$\boldsymbol{\alpha}_n\}$，令 $\boldsymbol{V}_2=\mathrm{span}\ \{\boldsymbol{\alpha}_{r+1},\ \cdots,\ \boldsymbol{\alpha}_n\}$，则 $\boldsymbol{V}_2$ 即为所求.

定义 1.4.4 若 $\boldsymbol{V}^n=\boldsymbol{V}_1\oplus\boldsymbol{V}_2$，则称 $\boldsymbol{V}_1$ 与 $\boldsymbol{V}_2$ 互补，$\boldsymbol{V}_1$ 与 $\boldsymbol{V}_2$ 分别称为 $\boldsymbol{V}_2$ 与 $\boldsymbol{V}_1$ 的补空间，并且称 $\boldsymbol{V}^n=\boldsymbol{V}_1\oplus\boldsymbol{V}_2$ 为直和分解，也称为空间分解.

直和的分解可以推广到有限多个子空间的情形

$$\mathbf{R}^n=\mathrm{Span}\{\boldsymbol{\varepsilon}_1\}\oplus\mathrm{Span}\{\boldsymbol{\varepsilon}_2\}\oplus\cdots\oplus\mathrm{Span}\{\boldsymbol{\varepsilon}_n\}$$

其中 $\boldsymbol{\varepsilon}_i=(0,\ \cdots,\ 0,\ \underset{(i)}{1},\ 0,\ \cdots,\ 0)$，$i=1,\ 2,\ \cdots,\ n$.

以下我们将这一部分做一个小结.

定理 1.4.12 设 $\boldsymbol{V}_1$，$\boldsymbol{V}_2$ 是线性空间 $\boldsymbol{V}$ 的任意两个子空间，$\boldsymbol{V}_1+\boldsymbol{V}_2$ 是直和的充分与必要条件是：

(1) $\boldsymbol{V}_1\cap\boldsymbol{V}_2=\mathbf{0}$；

(2) $\forall\boldsymbol{\alpha}\in\boldsymbol{V}_1+\boldsymbol{V}_2$，$\boldsymbol{\alpha}=\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2(\boldsymbol{\alpha}_1\in\boldsymbol{V}_1,\ \boldsymbol{\alpha}_2\in\boldsymbol{V}_2)$是唯一确定的；

(3) 零向量的分解式唯一，即若 $\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2=\mathbf{0}$，则必有 $\boldsymbol{\alpha}_1=\boldsymbol{\alpha}_2=\mathbf{0}(\boldsymbol{\alpha}_1\in\boldsymbol{V}_1,\ \boldsymbol{\alpha}_2\in\boldsymbol{V}_2)$；

(4) $\boldsymbol{V}^n$ 的基由 $\boldsymbol{V}_1$ 与 $\boldsymbol{V}_2$ 的基合并而成；

(5) $\dim(\boldsymbol{V}_1+\boldsymbol{V}_2)=\dim\boldsymbol{V}_1+\dim\boldsymbol{V}_2$.

【例 1.4.11】 已知 $\mathbf{R}^{n\times n}$的两个子空间

$$\boldsymbol{S}_1=\{\boldsymbol{A}\mid\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}\in\mathbf{R}^{n\times n}\},\boldsymbol{S}_2=\{\boldsymbol{A}\mid\boldsymbol{A}^{\mathrm{T}}=-\boldsymbol{A}\in\mathbf{R}^{n\times n}\}$$

证明：$\mathbf{R}^{n\times n}=\boldsymbol{S}_1\oplus\boldsymbol{S}_2$.

证 对任意 $\boldsymbol{A}\in\mathbf{R}^{n\times n}$，有

$$\boldsymbol{A}=\frac{\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}}{2}+\frac{\boldsymbol{A}-\boldsymbol{A}^{\mathrm{T}}}{2}$$

令

$$\boldsymbol{B}=\frac{\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}}{2},\boldsymbol{C}=\frac{\boldsymbol{A}-\boldsymbol{A}^{\mathrm{T}}}{2}$$

则有

$$\boldsymbol{B}^{\mathrm{T}}=\boldsymbol{B},\boldsymbol{C}^{\mathrm{T}}=-\boldsymbol{C}$$

从而 $\boldsymbol{B}\in\boldsymbol{S}_1$，$\boldsymbol{C}\in\boldsymbol{S}_2$，且 $\mathbf{R}^{n\times n}\subset\boldsymbol{S}_1+\boldsymbol{S}_2$.

若 $\boldsymbol{A}\in\boldsymbol{S}_1\cap\boldsymbol{S}_2$，$\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$，$\boldsymbol{A}^{\mathrm{T}}=-\boldsymbol{A}$，从而 $\boldsymbol{A}=0$，即 $\boldsymbol{S}_1\cap\boldsymbol{S}_2=\{\mathbf{0}\}$.
综上所证，有 $\mathbf{R}^{n\times n}=\boldsymbol{S}_1\oplus\boldsymbol{S}_2$.

§1.5 内 积 空 间

1. 内积的定义与性质

线性空间是解析几何中空间概念的推广，然而在线性空间中还缺少向量的度量的概念，例如向量的长度与夹角. 我们将在本节中引入这些重要的概念. 在 $\mathbf{R}^3$ 空间中，向量的长度与夹角是由向量的内积确定的. 我们先来回顾一下 $\mathbf{R}^3$ 中内积的概念，对 $\mathbf{R}^3$ 中任意两个非零向量 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$，它们的内积 $\boldsymbol{\alpha}\cdot\boldsymbol{\beta}=|\boldsymbol{\alpha}||\boldsymbol{\beta}|\cos\theta$，其中 $|\boldsymbol{\alpha}|$，$|\boldsymbol{\beta}|$ 分别是 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$ 的长度，θ 是 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的夹角. 并且当 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$ 中有一个是零向量时，$\boldsymbol{\alpha}\cdot\boldsymbol{\beta}=0$.

有了内积的概念后，$\mathbf{R}^3$ 中任何一个向量 $\boldsymbol{\alpha}$ 的长度就可以由式 $|\boldsymbol{\alpha}|=\sqrt{\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}}$ 确定，同时两个向量 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 之间的夹角 θ 也可以由式 $\cos\theta=\dfrac{\boldsymbol{\alpha}\cdot\boldsymbol{\beta}}{|\boldsymbol{\alpha}||\boldsymbol{\beta}|}$ 确定.

当然不能直接将 $\mathbf{R}^3$ 中的内积公式推广到一般的线性空间，与定义线性空间类似，我们用公理来引入一般线性空间的内积.

在 $\mathbf{R}^3$ 中，向量的内积具有以下性质（$\forall \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbf{R}^3$，$\forall a \in \mathbf{R}$）：

(1) 对称性：$\boldsymbol{\alpha} \cdot \boldsymbol{\beta} = \boldsymbol{\beta} \cdot \boldsymbol{\alpha}$；

(2) 线性性：$(\boldsymbol{\alpha}+\boldsymbol{\beta}) \cdot \boldsymbol{\gamma} = \boldsymbol{\alpha} \cdot \boldsymbol{\gamma} + \boldsymbol{\beta} \cdot \boldsymbol{\gamma}$，

$(a\boldsymbol{\alpha}) \cdot \boldsymbol{\beta} = a(\boldsymbol{\alpha} \cdot \boldsymbol{\beta})$；

(3) 正定性：当 $\boldsymbol{\alpha} \neq \mathbf{0}$ 时，$\boldsymbol{\alpha} \cdot \boldsymbol{\alpha} > 0$.

这些性质足以刻画内积的概念.

定义 1.5.1 设 $\mathbf{V}$ 是实数域 $\mathbf{R}$ 上的线性空间，$\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbf{V}$；$a, b \in \mathbf{R}$. 在 $\mathbf{V}$ 中定义了一个实数 $(\boldsymbol{\alpha}, \boldsymbol{\beta})$，它满足下列条件：

(1) 对称性：$(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (\boldsymbol{\beta}, \boldsymbol{\alpha})$；

(2) 线性性：$(\boldsymbol{\alpha}+\boldsymbol{\beta}, \boldsymbol{\gamma}) = (\boldsymbol{\alpha}, \boldsymbol{\gamma}) + (\boldsymbol{\beta}, \boldsymbol{\gamma})$，

$(a\boldsymbol{\alpha}, \boldsymbol{\beta}) = a(\boldsymbol{\alpha}, \boldsymbol{\beta})$；

(3) 正定性：$(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \geqslant \mathbf{0}$，当且仅当 $\boldsymbol{\alpha} = \mathbf{0}$ 时，$(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = 0$.

则称实数 $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ 为向量 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的内积，称定义了内积的线性空间 $\mathbf{V}$ 为 **Euclid** 空间，简称为欧氏空间. 也称 $\mathbf{V}$ 为实内积空间.

显然，欧氏空间的子空间也是欧氏空间.

【例 1.5.1】 考虑线性空间 $\mathbf{R}^n$，对任意的 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbf{R}^n$，不妨设 $\boldsymbol{\alpha} = (a_1, a_2, \cdots, a_n)$，$\boldsymbol{\beta} = (b_1, b_2, \cdots, b_n)$，规定 $(\boldsymbol{\alpha}, \boldsymbol{\beta}) = a_1b_1 + a_2b_2 + \cdots + a_nb_n$，满足定义 1.5.1 的条件，所以线性空间 $\mathbf{R}^n$ 对于如上规定的运算构成一个欧氏空间.

【例 1.5.2】 考虑线性空间 $\mathbf{R}^n$，对任意的 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbf{R}^n$，不妨设 $\boldsymbol{\alpha} = (a_1, a_2, \cdots, a_n)$，$\boldsymbol{\beta} = (b_1, b_2, \cdots, b_n)$，规定 $(\boldsymbol{\alpha}, \boldsymbol{\beta}) = a_1b_1 + 2a_2b_2 + \cdots + na_nb_n$，不难验证线性空间 $\mathbf{R}^n$ 对于如上规定的运算也构成一个欧氏空间.

由此可见，对于同一个线性空间可以引入不同的内积，从而构成不同的欧氏空间.

【例 1.5.3】 设 $\mathbf{C}[a, b]$ 为定义在区间 $[a, b]$ 上一切连续实函数所构成的线性空间，对任意的 $f(x), g(x), h(x) \in \mathbf{C}[a, b]$，规定 $(f,g) = \int_a^b f(x)g(x)\mathrm{d}x$.

$\because f(x)g(x) \in \mathbf{C}[a, b]$，$\therefore (f, g)$ 是唯一确定的实数. 并且：

(1) $(f,g) = \int_a^b f(x)g(x)\mathrm{d}x = \int_a^b g(x)f(x)\mathrm{d}x = (g,f)$；

(2) $(f+g,h) = \int_a^b (f(x)+g(x))h(x)\mathrm{d}x = \int_a^b f(x)h(x)\mathrm{d}x + \int_a^b g(x)h(x)\mathrm{d}x$

$= (f,h) + (g,h)$，

$(kf,g) = \int_a^b (kf(x))g(x)\mathrm{d}x = k\int_a^b f(x)g(x)\mathrm{d}x = k(f,g)$；

(3) $(f,f) = \int_a^b f^2(x)\mathrm{d}x \geqslant 0$，当且仅当 $f(x) = 0$ 时，$(f, f) = 0$.

所以线性空间 $\mathbf{C}[a, b]$ 是一个欧氏空间.

【例 1.5.4】 设 H 是一切平方和收敛的实数列 $\boldsymbol{\xi} = (x_1, x_2, \cdots, x_n, \cdots)$，$\sum_{n=1}^{\infty} x_n^2 < +\infty$ 所成的集合.

在 H 中规定 $\xi+\eta=(x_1+y_1, x_2+y_2, \cdots, x_n+y_n, \cdots)$，$a\xi=(ax_1, ax_2, \cdots, ax_n, \cdots)$；内积 $(\boldsymbol{\xi},\boldsymbol{\eta})=\sum\limits_{n=1}^{\infty} x_n y_n$. 试验证：$H$ 是一个欧氏空间（其中：$\boldsymbol{\xi}=(x_1, x_2, \cdots, x_n, \cdots)$，$\boldsymbol{\eta}=(y_1, y_2, \cdots, y_n, \cdots)\in H$，$a\in\mathbf{R}$）.

证 首先验证以上定义的三种运算的合理性.

$\because |x_n y_n|\leqslant\frac{1}{2}(x_n^2+y_n^2)$，$\therefore \sum\limits_{n=1}^{\infty} x_n y_n$ 收敛；

又 $\because \sum\limits_{k=n}^{n+m}(ax_k)^2=a^2\sum\limits_{k=n}^{n+m}x_k^2, \sum\limits_{k=n}^{n+m}(x_k+y_k)^2=\sum\limits_{k=n}^{n+m}x_k^2+2\sum\limits_{k=n}^{n+m}x_k y_k+\sum\limits_{k=n}^{n+m}y_k^2$，

$\therefore \sum\limits_{n=1}^{\infty}(ax_n)^2$ 与 $\sum\limits_{n=1}^{\infty}(x_n+y_n)^2$ 收敛.

因此，对于任意的 $\boldsymbol{\xi}=(x_1, x_2, \cdots, x_n, \cdots)$，$\boldsymbol{\eta}=(y_1, y_2, \cdots, y_n, \cdots)\in H$，$a\in\mathbf{R}$ 有 $\boldsymbol{\xi}+\boldsymbol{\eta}\in H$，$a\xi\in H$. 其次验证线性空间 H 为一个欧氏空间（略）. 空间 H 称为**Hilbert** 空间.

定义 1.5.2 在欧氏空间 $\mathbf{V}$ 中，非负实数 $\sqrt{(\boldsymbol{\alpha}, \boldsymbol{\alpha})}$ 称为 $\mathbf{V}$ 中向量 $\boldsymbol{\alpha}$ 的长度（模或范数），记为 $|\boldsymbol{\alpha}|=\sqrt{(\boldsymbol{\alpha}, \boldsymbol{\alpha})}$（或 $\|\boldsymbol{\alpha}\|$）.

由此可知，当 $\boldsymbol{\alpha}\neq\mathbf{0}$ 时，$|\boldsymbol{\alpha}|>0$；$|\boldsymbol{\alpha}|=0$ 当且仅当 $\boldsymbol{\alpha}=\mathbf{0}$.

【例 1.5.5】 在欧氏空间 $\mathbf{R}^n$ 中，对任意的向量 $\boldsymbol{\alpha}=(a_1, a_2, \cdots, a_n)\in\mathbf{R}^n$，有 $|\boldsymbol{\alpha}|=\sqrt{a_1^2+a_2^2+\cdots+a_n^2}$.

在欧氏空间 $\mathbf{C}[a, b]$ 中，对任意的向量 $f(x)\in\mathbf{C}[a, b]$，有 $|f(x)|=\sqrt{\int_a^b f^2(x)\mathrm{d}x}$.

在 **Hilbert** 空间 H 中，对任意的向量 $\boldsymbol{\xi}\in H$，有 $|\boldsymbol{\xi}|=\sqrt{\sum\limits_{n=1}^{\infty}x_n^2}$.

长度为 1 的向量叫做单位向量. 如果 $\boldsymbol{\xi}\neq\mathbf{0}$，用向量 $\boldsymbol{\xi}$ 的模 $|\boldsymbol{\xi}|$ 的倒数乘以 $\boldsymbol{\xi}$，则得到一个与 $\boldsymbol{\xi}$ 同方向的单位向量 $\frac{\boldsymbol{\xi}}{|\boldsymbol{\xi}|}$，这一运算我们称之为将向量 $\boldsymbol{\xi}$ 单位化或规范化.

定义 1.5.3 在欧氏空间中非零向量 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$ 的夹角由式 $\theta=\arccos\frac{(\boldsymbol{\alpha}, \boldsymbol{\beta})}{|\boldsymbol{\alpha}||\boldsymbol{\beta}|}$ 确定，其中 $0\leqslant\theta\leqslant\pi$.

向量的内积有以下诸性质：

性质 1 $(\boldsymbol{\alpha}, k\boldsymbol{\beta})=k(\boldsymbol{\alpha}, \boldsymbol{\beta})$；

性质 2 $(\boldsymbol{\alpha}, \mathbf{0})=(\mathbf{0}, \boldsymbol{\alpha})=0$；

性质 3 $\left(\sum\limits_{i=1}^{n}a_i x_i, \sum\limits_{j=1}^{n}b_j y_j\right)=\sum\limits_{i,j=1}^{n}a_i b_j(x_i, y_j)$.

其中 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$ 是向量，a_i，b_j，x_i，y_j 是数量，并且 $\boldsymbol{\alpha}=\sum\limits_{i=1}^{n}a_i x_i, \boldsymbol{\beta}=\sum\limits_{j=1}^{n}b_j y_j$.

欧氏空间是对实数域 $\mathbf{R}$ 上的线性空间而言，对于复数域 $\mathbf{C}$ 上的线性空间我们有以下定义.

定义 1.5.4 设 $\mathbf{V}$ 是复数域 $\mathbf{C}$ 上的线性空间，对于 $\mathbf{V}$ 中任意两个向量 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$，依照某一规则有一个复数 $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ 与之对应，并且满足下列三个条件：

(1) 对称性：$(\boldsymbol{\alpha}, \boldsymbol{\beta})=\overline{(\boldsymbol{\beta}, \boldsymbol{\alpha})}$，其中$\overline{(\boldsymbol{\beta}, \boldsymbol{\alpha})}$是$(\boldsymbol{\beta}, \boldsymbol{\alpha})$的共轭复数；

(2) 线性性：$(\boldsymbol{\alpha}+\boldsymbol{\beta}, \boldsymbol{\gamma})=(\boldsymbol{\alpha}, \boldsymbol{\gamma})+(\boldsymbol{\beta}, \boldsymbol{\gamma})$，

$(a\boldsymbol{\alpha},\boldsymbol{\beta})=a(\boldsymbol{\alpha},\boldsymbol{\beta}), \forall a \in \mathbf{C}$；

(3) 正定性：$(\boldsymbol{\alpha}, \boldsymbol{\alpha})\geqslant 0$，当且仅当$\boldsymbol{\alpha}=\mathbf{0}$时，实数$(\boldsymbol{\alpha}, \boldsymbol{\alpha})=0$.

则称复数$(\boldsymbol{\alpha}, \boldsymbol{\beta})$为向量$\boldsymbol{\alpha}$与$\boldsymbol{\beta}$的内积，称定义了内积的复线性空间$\mathbf{V}$为酉空间(Unitary Space)，也称$\mathbf{V}$为复内积空间.

定义 1.5.5 在酉空间$\mathbf{V}$中，非负实数$\sqrt{(\boldsymbol{\alpha}, \boldsymbol{\alpha})}$称为$\mathbf{V}$中向量$\alpha$的长度（模或范数），记为$|\boldsymbol{\alpha}|=\sqrt{(\boldsymbol{\alpha}, \boldsymbol{\alpha})}$（或$\|\boldsymbol{\alpha}\|$）.

两个非零向量$\boldsymbol{\alpha}$，$\boldsymbol{\beta}$的夹角由式$\cos^2\theta=\dfrac{(\boldsymbol{\alpha}, \boldsymbol{\beta})(\boldsymbol{\beta}, \boldsymbol{\alpha})}{(\boldsymbol{\alpha}, \boldsymbol{\alpha})(\boldsymbol{\beta}, \boldsymbol{\beta})}$确定.

对酉空间$\mathbf{V}$而言，前面的三个性质对应为：

性质 1 $(\boldsymbol{\alpha}, k\boldsymbol{\beta})=\bar{k}(\boldsymbol{\alpha}, \boldsymbol{\beta})$；

性质 2 $(\boldsymbol{\alpha}, \mathbf{0})=(\mathbf{0}, \boldsymbol{\alpha})=0$；

性质 3 $\left(\sum\limits_{i=1}^{n}a_ix_i, \sum\limits_{j=1}^{n}b_jy_j\right)=\sum\limits_{i,j=1}^{n}a_i\overline{b_j}(x_i,y_j)$.

在线性空间中确定了向量的长度与向量的夹角之后，就可以对向量的正交性进行讨论了.

2. 向量的正交性与施米特（Schmidt）正交化方法

定义 1.5.6 设$\mathbf{V}^n$为n维欧氏空间（或酉空间），对任意的向量$\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\mathbf{V}^n$，如果$(\boldsymbol{\alpha}, \boldsymbol{\beta})=0$，则称$\boldsymbol{\alpha}$与$\boldsymbol{\beta}$正交，记作$\boldsymbol{\alpha}\perp\boldsymbol{\beta}$.

【例 1.5.6】 在$\mathbf{R}^n$中，向量组$\{\boldsymbol{\varepsilon}_i \mid \boldsymbol{\varepsilon}_i=(0, \cdots, 0, \underset{(i)}{1}, 0, \cdots, 0), i=1, 2, \cdots, n\}$中任意两个向量$\boldsymbol{\varepsilon}_i$，$\boldsymbol{\varepsilon}_j(i\neq j)$正交.

【例 1.5.7】 零向量与任意向量正交.

定理 1.5.1 若$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_k$是$\mathbf{R}^n$（或$\mathbf{C}^n$）中一个非零正交系，即$\boldsymbol{\alpha}_i\neq\mathbf{0}$，$(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j)=0$，$(i\neq j; i, j=1, 2, \cdots, k)$，则$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_k$必线性无关.

证 设$k_i\in\mathbf{R}$（或$\mathbf{C}$）$(i=1, 2, \cdots, k)$，考虑$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_k$的线性组合$k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_k\boldsymbol{\alpha}_k=0$，对于任意的$\boldsymbol{\alpha}_j$，必有$\left(\sum\limits_{i=1}^{k}k_i\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_j\right)=\sum\limits_{i=1}^{k}k_i(\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_j)=k_j(\boldsymbol{\alpha}_j,\boldsymbol{\alpha}_j)=0$，由于$\boldsymbol{\alpha}_j\neq 0$，$(\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_j)>0$，则$k_j=0(j=1, 2, \cdots, k)$，所以$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_k$线性无关.

定义 1.5.7 设$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_k$为$\mathbf{R}^n$（或$\mathbf{C}^n$）中一个正交系，如果$|\boldsymbol{\alpha}_i|=\sqrt{(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_i)}=1$ $(i=1, 2, \cdots, k)$，则称$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_k$为$\mathbf{R}^n$（或$\mathbf{C}^n$）中一个标准正交系.

定理 1.5.2 在空间$\mathbf{C}^n$（或$\mathbf{R}^n$）中必存在标准正交基$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$，即

$$(\boldsymbol{\alpha}_j,\boldsymbol{\alpha}_j)=\delta_{ij}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases}(i,j=1,2,\cdots,n)$$

符号δ_{ij}叫做 Kronecker 符号.

证 设$\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，…，$\boldsymbol{\beta}_n$是空间$\mathbf{C}^n$（或$\mathbf{R}^n$）中的一个基，首先取$\boldsymbol{\gamma}_1=\boldsymbol{\beta}_1$，则$\boldsymbol{\gamma}_1$是$\boldsymbol{\beta}_1$的线性组合，并且$\boldsymbol{\gamma}_1\neq\mathbf{0}$；其次取$\boldsymbol{\gamma}_2=\boldsymbol{\beta}_2-\dfrac{(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_1)}{(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_1)}\boldsymbol{\gamma}_1$，则$\boldsymbol{\gamma}_2$是$\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$的线性组合，因为$\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$线性无关，所以$\boldsymbol{\gamma}_2\neq\mathbf{0}$.

又因为 $(\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_1)=(\boldsymbol{\beta}_2,\boldsymbol{\gamma}_1)-\dfrac{(\boldsymbol{\beta}_2,\boldsymbol{\gamma}_1)}{(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_1)}(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_1)=0$，所以 $\boldsymbol{\gamma}_2$ 与 $\boldsymbol{\gamma}_1$ 正交.

假设 $1<k\leqslant n$，而满足条件的 $\boldsymbol{\gamma}_1$，$\boldsymbol{\gamma}_2$，…，$\boldsymbol{\gamma}_{k-1}$ 都已作出.

取 $\boldsymbol{\gamma}_k=l_1\boldsymbol{\beta}_1+l_2\boldsymbol{\beta}_2+\cdots+l_{k-1}\boldsymbol{\beta}_{k-1}+\boldsymbol{\beta}_k$，因此 $\boldsymbol{\gamma}_k$ 是 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，…，$\boldsymbol{\beta}_k$ 的线性组合. 由 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，…，$\boldsymbol{\beta}_{k-1}$，$\boldsymbol{\beta}_k$ 线性无关，可知 $\boldsymbol{\gamma}_k\neq\mathbf{0}$，又因为 $\boldsymbol{\gamma}_1$，$\boldsymbol{\gamma}_2$，…，$\boldsymbol{\gamma}_{k-1}$ 两两正交，所以 $(\boldsymbol{\gamma}_k,\ \boldsymbol{\gamma}_i)=(\boldsymbol{\beta}_k,\ \boldsymbol{\gamma}_i)-\dfrac{(\boldsymbol{\beta}_k,\ \boldsymbol{\gamma}_i)}{(\boldsymbol{\gamma}_i,\ \boldsymbol{\gamma}_i)}(\boldsymbol{\gamma}_i,\ \boldsymbol{\gamma}_i)=0$， $i=1,\ 2,\ \cdots,\ k-1$.

因此 $\boldsymbol{\gamma}_1$，$\boldsymbol{\gamma}_2$，…，$\boldsymbol{\gamma}_k$ 满足定理的要求.

如此可以由空间 $\mathbf{C}^n$（或 $\mathbf{R}^n$）的一个基 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，…，$\boldsymbol{\beta}_n$，通过上面的方法得到空间 $\mathbf{C}^n$（或 $\mathbf{R}^n$）的一个正交基 $\boldsymbol{\gamma}_1$，$\boldsymbol{\gamma}_2$，…，$\boldsymbol{\gamma}_n$. 再令 $\boldsymbol{\alpha}_i=\dfrac{\boldsymbol{\gamma}_i}{|\boldsymbol{\gamma}_i|}$，$i=1,\ 2,\ \cdots,\ n$. 则 $\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n\}$ 就是空间 $\mathbf{C}^n$（或 $\mathbf{R}^n$）的一个标准正交基. 这一方法叫做施米特（**Schmidt**）正交化方法.

【例 1.5.8】 在欧氏空间 $\mathbf{R}^3$ 中对于基 $\boldsymbol{\beta}_1=(1,\ 1,\ 1)$，$\boldsymbol{\beta}_2=(1,\ 1,\ 0)$，$\boldsymbol{\beta}_3=(1,\ 0,\ 0)$ 施行正交化方法，求出 $\mathbf{R}^3$ 的一个标准正交基.

取 $\boldsymbol{\gamma}_1=\boldsymbol{\beta}_1=(1,\ 1,\ 1)$，由施米特正交化方法

$$\boldsymbol{\gamma}_2=\boldsymbol{\beta}_2-\frac{(\boldsymbol{\beta}_2,\boldsymbol{\gamma}_1)}{(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_1)}\boldsymbol{\gamma}_1=\left(\frac{1}{3},\frac{1}{3},-\frac{2}{3}\right)$$

$$\boldsymbol{\gamma}_3=\boldsymbol{\beta}_3-\frac{(\boldsymbol{\beta}_3,\boldsymbol{\gamma}_1)}{(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_1)}\boldsymbol{\gamma}_1-\frac{(\boldsymbol{\beta}_3,\boldsymbol{\gamma}_2)}{(\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_2)}\boldsymbol{\gamma}_2=\left(\frac{1}{2},-\frac{1}{2},0\right)$$

所以 $\{\boldsymbol{\gamma}_1,\ \boldsymbol{\gamma}_2,\ \boldsymbol{\gamma}_3\}$ 是 $\mathbf{R}^3$ 的一个正交基；
再令

$$\boldsymbol{\alpha}_1=\frac{\boldsymbol{\gamma}_1}{|\boldsymbol{\gamma}_1|}=\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)$$

$$\boldsymbol{\alpha}_2=\frac{\boldsymbol{\gamma}_2}{|\boldsymbol{\gamma}_2|}=\left(\frac{1}{\sqrt{6}},\frac{1}{\sqrt{6}},-\frac{2}{\sqrt{6}}\right)$$

$$\boldsymbol{\alpha}_3=\frac{\boldsymbol{\gamma}_3}{|\boldsymbol{\gamma}_3|}=\left(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}},0\right)$$

则 $\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \boldsymbol{\alpha}_3\}$ 即为欧氏空间 $\mathbf{R}^3$ 的一个标准正交基.

【例 1.5.9】 在 $\boldsymbol{P}_3[t]$ 中按内积 $(f(t),g(t))=\int_{-1}^{1}f(t)g(t)\mathrm{d}t,(\forall f(t),g(t)\in\boldsymbol{P}_3[t])$ 做成的欧氏空间中，求一组标准正交基.

解 取 $\boldsymbol{P}_3[t]$ 的基 $1,\ t,\ t^2,\ t^3$，将其正交化得

$$g_1(t)=1$$

$$g_2(t)=t-\frac{(t,g_1)}{(g_1,g_1)}g_1=t-\frac{\int_{-1}^{1}t\mathrm{d}t}{\int_{-1}^{1}1\mathrm{d}t}\times1=t$$

$$g_3(t)=t^2-\frac{(t,g_1)}{(g_1,g_1)}g_2-\frac{(t^2,g_2)}{(g_2,g_2)}g_2=t^2-\frac{\int_{-1}^{1}t^2\mathrm{d}t}{\int_{-1}^{1}1\mathrm{d}t}\times1-\frac{\int_{-1}^{1}t^3\mathrm{d}t}{\int_{-1}^{1}t^2\mathrm{d}t}t=t^2-\frac{1}{3}$$

$$g_4(t)=t^3-\frac{(t^3,g_1)}{(g_1,g_1)}g_1-\frac{(t^3,g_2)}{(g_2,g_2)}g_2-\frac{(t^3,g_3)}{(g_3,g_3)}g_3$$

$$=t^3-\frac{\int_{-1}^{1}t^3\mathrm{d}t}{\int_{-1}^{1}1\mathrm{d}t}\times 1-\frac{\int_{-1}^{1}t^4\mathrm{d}t}{\int_{-1}^{1}t^2\mathrm{d}t}t-\frac{\int_{-1}^{1}t^3\left(t^2-\frac{1}{3}\right)\mathrm{d}t}{\int_{-1}^{1}\left(t^2-\frac{1}{3}\right)\mathrm{d}t}\left(t^2-\frac{1}{3}\right)=t^3-\frac{3}{5}t$$

再单位化即可得到 $P_3[t]$ 的一组标准正交基

$$h_1=\frac{1}{|g_1(t)|}g_1(t)=\frac{1}{\sqrt{\int_{-1}^{1}1\mathrm{d}t}}\times 1=\frac{1}{\sqrt{2}}$$

$$h_2=\frac{1}{|g_2(t)|}g_2(t)=\frac{1}{\sqrt{\int_{-1}^{1}t^2\mathrm{d}t}}t=\frac{1}{\sqrt{\frac{2}{3}}}t=\frac{\sqrt{6}}{2}t$$

$$h_3=\frac{1}{|g_3(t)|}g_3(t)=\frac{1}{\sqrt{\int_{-1}^{1}\left(t^2-\frac{1}{3}\right)^2\mathrm{d}t}}\left(t^2-\frac{1}{3}\right)=\frac{1}{\sqrt{\frac{8}{45}}}\left(t^2-\frac{1}{3}\right)=\frac{\sqrt{10}}{4}(3t^2-1)$$

$$h_4=\frac{1}{|g_4(t)|}g_4(t)=\frac{1}{\sqrt{\int_{-1}^{1}\left(t^3-\frac{3}{5}t\right)^2\mathrm{d}t}}\left(t^3-\frac{3}{5}t\right)=\frac{1}{\sqrt{\frac{8}{175}}}\left(t^3-\frac{3}{5}t\right)=\frac{\sqrt{14}}{4}(5t^3-3t)$$

由于对 $\mathbf{R}^3$ 而言 $\boldsymbol{\varepsilon}_1=(1,0,0)$，$\boldsymbol{\varepsilon}_2=(0,1,0)$，$\boldsymbol{\varepsilon}_3=(0,0,1)$也是一个标准正交基. 因此可以看出欧氏空间中的标准正交基不唯一.

定理 1.5.3 设 e_1，e_2，…，e_n 是酉空间 $\mathbf{C}^n$ 中的一个标准正交基，对任意的向量 $\boldsymbol{\alpha}\in\mathbf{C}^n$，$\boldsymbol{\alpha}$ 在基 e_1，e_2，…，e_n 下的坐标为$(x_1, x_2, \cdots, x_n)$，则 $x_i=(\boldsymbol{\alpha}, e_i)$ $(i=1,2,\cdots,n)$.

证 因为 $\boldsymbol{\alpha}=\sum_{j=1}^{n}x_je_j$，所以 $(\boldsymbol{\alpha},e_i)=\left(\sum_{j=1}^{n}x_je_j,e_i\right)=\sum_{j=1}^{n}x_j(e_j,e_i)=x_i(e_i,e_i)=x_i$ $(i=1,2,\cdots,n)$.

定理 1.5.4 设 $\{e_1, e_2, \cdots, e_n\}$ 与 $\{e'_1, e'_2, \cdots, e'_n\}$ 是酉空间 $\mathbf{C}^n$ 中的两个标准正交基，则 $\{e_1, e_2, \cdots, e_n\}$ 到 $\{e'_1, e'_2, \cdots, e'_n\}$ 的过渡矩阵为酉矩阵.

证 设 $e'_i=\sum_{j=1}^{n}c_{ji}e_j$，则 $\delta_{ij}=(e'_i,e'_j)=\left(\sum_{l=1}^{n}c_{li}e_l,\sum_{k=1}^{n}c_{kj}e_k\right)$

$$=\sum_{l=1}^{n}c_{li}\left(e_l,\sum_{k=1}^{n}c_{kj}e_k\right)=\sum_{l=1}^{n}c_{li}\overline{\left(\sum_{k=1}^{n}c_{kj}e_k,e_l\right)}=\sum_{l=1}^{n}c_{li}\sum_{k=1}^{n}\overline{c_{kj}}\,\overline{(e_k,e_l)}$$

$$=\sum_{l=1}^{n}c_{li}\sum_{k=1}^{n}\overline{c_{kj}}(e_l,e_k)=\sum_{l=1}^{n}\sum_{k=1}^{n}c_{li}\overline{c_{kj}}\delta_{lk}=\sum_{l=1}^{n}\sum_{k=1}^{n}c_{li}\overline{c_{lj}}$$

∴ $\overline{U}^{\mathrm{T}}U=E$ 或 $U^*U=E$，即矩阵 U 为酉矩阵 $(U^*=\overline{U}^{\mathrm{T}})$

3. 内积的表示与 Cauchy-Schwarz 不等式

在线性空间 $\mathbf{V}$ 中，取一个基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_r$，对于任意向量 $\boldsymbol{\alpha}\in V$，$\boldsymbol{\alpha}$ 可以表示为 $\boldsymbol{\alpha}=\sum_{i=1}^{r}a_i\boldsymbol{\alpha}_i$. 由此向量 $\boldsymbol{\alpha}$ 与它的坐标 $(a_1, a_2, \cdots, a_r)$ 之间就建立了一个一一对应的关系，因此向量的内积也可以用它们的坐标来表示.

设酉空间 $\mathbf{C}^n$ 中的一个基为 $\boldsymbol{\eta}_1$，$\boldsymbol{\eta}_2$，…，$\boldsymbol{\eta}_n$，这个基不一定正交. 对 $\mathbf{C}^n$ 中的任意两个向

量$\boldsymbol{\alpha}$，$\boldsymbol{\beta}$，有$\boldsymbol{\alpha}=\sum_{i=1}^{n}x_i\boldsymbol{\eta}_i,\boldsymbol{\beta}=\sum_{j=1}^{n}y_j\boldsymbol{\eta}_j$.

由内积的定义可知$(\boldsymbol{\alpha},\boldsymbol{\beta})=(\sum_{i=1}^{n}x_i\boldsymbol{\eta}_i,\sum_{j=1}^{n}y_j\boldsymbol{\eta}_j)=\sum_{i,j=1}^{n}x_i\overline{y_j}(\boldsymbol{\eta}_i,\boldsymbol{\eta}_j)=\sum_{i,j=1}^{n}h_{ij}x_iy_j$.

其中$h_{ij}=(\boldsymbol{\eta}_i,\boldsymbol{\eta}_j)$，$i$，$j=1$，2，…，$n$. 又因为$(\boldsymbol{\eta}_i,\boldsymbol{\eta}_j)=\overline{(\boldsymbol{\eta}_j,\boldsymbol{\eta}_i)}$，所以$h_{ij}=\overline{h_{ji}}$；而$h_{ij}$只与基有关，与$\boldsymbol{\alpha}$，$\boldsymbol{\beta}$无关，∴$(\boldsymbol{\alpha},\boldsymbol{\beta})=y^*\boldsymbol{H}x$，其中$\boldsymbol{H}=(h_{ij})_{n\times n}$，$x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，$y=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$，$y^*=\overline{y}^{\mathrm{T}}$；并且$\boldsymbol{H}=\boldsymbol{H}^*$，称$\boldsymbol{H}$为 **Hermite** 矩阵.

由内积的定义知$(\boldsymbol{\alpha},\boldsymbol{\alpha})\geqslant0$，当且仅当$\boldsymbol{\alpha}=\boldsymbol{0}$时，才有$(\boldsymbol{\alpha},\boldsymbol{\alpha})=0$. 而$(\boldsymbol{\alpha},\boldsymbol{\alpha})=x^*\boldsymbol{H}x$，所以$\boldsymbol{H}$是一个 **Hermite** 矩阵.

同理可知，欧氏空间的内积可以表示为$(\boldsymbol{\alpha},\boldsymbol{\beta})=y^{\mathrm{T}}\boldsymbol{H}x$，其中$\boldsymbol{H}$是一个正定矩阵. 如果所取的基是标准正交基，从以上的讨论可知有$(\boldsymbol{\alpha},\boldsymbol{\beta})=\sum_{i=1}^{n}x_iy_i$. 并且：

当$\boldsymbol{\alpha}\in\mathbf{C}^n$时 $$|\boldsymbol{\alpha}|^2=(\sqrt{(\boldsymbol{\alpha},\boldsymbol{\alpha})})^2=(\boldsymbol{\alpha},\boldsymbol{\alpha})=\sum_{i=1}^{n}|x_i|^2$$

当$\boldsymbol{\alpha}\in\mathbf{R}^n$时 $$|\boldsymbol{\alpha}|^2=\sum_{i=1}^{n}x_i^2$$

定理 1.5.5 （**Cauchy-Schwarz** 不等式）对任意的向量$\boldsymbol{\alpha}$，$\boldsymbol{\beta}$属于酉空间$\mathbf{C}^n$，下列不等式恒成立$|(\boldsymbol{\alpha},\boldsymbol{\beta})|\leqslant|\boldsymbol{\alpha}||\boldsymbol{\beta}|$.

证 若$\boldsymbol{\beta}\neq0$，则$\forall\lambda\in\mathbf{R}$，有$(\boldsymbol{\alpha}-\lambda\boldsymbol{\beta},\alpha-\lambda\boldsymbol{\beta})=(\boldsymbol{\alpha},\boldsymbol{\alpha})-\overline{\lambda}(\boldsymbol{\alpha},\boldsymbol{\beta})-\lambda(\boldsymbol{\beta},\boldsymbol{\alpha})+\lambda\overline{\lambda}(\boldsymbol{\beta},\boldsymbol{\beta})\geqslant0$，取$\lambda=\dfrac{(\boldsymbol{\alpha},\boldsymbol{\beta})}{(\boldsymbol{\beta},\boldsymbol{\beta})}$，以正数$(\boldsymbol{\beta},\boldsymbol{\beta})$乘以上式两侧，整理有

$$(\boldsymbol{\alpha},\boldsymbol{\alpha})(\boldsymbol{\beta},\boldsymbol{\beta})-\overline{(\boldsymbol{\alpha},\boldsymbol{\beta})}(\boldsymbol{\alpha},\boldsymbol{\beta})-(\boldsymbol{\alpha},\boldsymbol{\beta})(\boldsymbol{\beta},\boldsymbol{\alpha})+(\boldsymbol{\alpha},\boldsymbol{\beta})\overline{(\boldsymbol{\alpha},\boldsymbol{\beta})}\geqslant0$$

即 $$(\boldsymbol{\alpha},\boldsymbol{\alpha})(\boldsymbol{\beta},\boldsymbol{\beta})-2(\boldsymbol{\alpha},\boldsymbol{\beta})\overline{(\boldsymbol{\alpha},\boldsymbol{\beta})}+(\boldsymbol{\alpha},\boldsymbol{\beta})\overline{(\boldsymbol{\alpha},\boldsymbol{\beta})}\geqslant0,$$

$$(\boldsymbol{\alpha},\boldsymbol{\beta})\overline{(\boldsymbol{\alpha},\boldsymbol{\beta})}\leqslant(\boldsymbol{\alpha},\boldsymbol{\alpha})(\boldsymbol{\beta},\boldsymbol{\beta})=|\boldsymbol{\alpha}|^2|\boldsymbol{\beta}|^2$$

两侧开方取算术根，得到$|(\boldsymbol{\alpha},\boldsymbol{\beta})|\leqslant|\boldsymbol{\alpha}||\boldsymbol{\beta}|$；

若$\boldsymbol{\alpha}$，$\boldsymbol{\beta}$线性相关，不妨设$\boldsymbol{\alpha}=k\boldsymbol{\beta}$，则$|(\boldsymbol{\alpha},\boldsymbol{\beta})|=|k(\boldsymbol{\beta},\boldsymbol{\beta})|=|k||\boldsymbol{\beta}|^2=|k\boldsymbol{\beta}||\boldsymbol{\beta}|=|\boldsymbol{\alpha}||\boldsymbol{\beta}|$；

若$\boldsymbol{\alpha}$，$\boldsymbol{\beta}$线性无关，则必有$\boldsymbol{\alpha}-\lambda\boldsymbol{\beta}\neq\boldsymbol{0}$，则$(\boldsymbol{\alpha}-\lambda\boldsymbol{\beta},\boldsymbol{\alpha}-\lambda\boldsymbol{\beta})>0$

所以$|(\boldsymbol{\alpha},\boldsymbol{\beta})|<|\boldsymbol{\alpha}||\boldsymbol{\beta}|$.

由 **Cauchy-Schwarz** 不等式可得

(1) $|\boldsymbol{\alpha}+\boldsymbol{\beta}|\leqslant|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|$；

(2) $|\boldsymbol{\alpha}-\boldsymbol{\beta}|\geqslant\big||\boldsymbol{\alpha}|-|\boldsymbol{\beta}|\big|$.

我们称$|\boldsymbol{\alpha}-\boldsymbol{\beta}|$为向量$\boldsymbol{\alpha}$与$\boldsymbol{\beta}$之间的距离，记为$\rho(\boldsymbol{\alpha},\boldsymbol{\beta})=|\boldsymbol{\alpha}-\boldsymbol{\beta}|$. 它有以下性质：

性质 1 $\rho(\boldsymbol{\alpha},\boldsymbol{\alpha})=0$，$\rho(\boldsymbol{\alpha},\boldsymbol{\beta})>0$；

性质 2 $\rho(\boldsymbol{\alpha},\boldsymbol{\beta})=\rho(\boldsymbol{\beta},\boldsymbol{\alpha})$；

性质 3 $\rho(\boldsymbol{\alpha},\boldsymbol{\beta})+\rho(\boldsymbol{\beta},\boldsymbol{\gamma})\geqslant\rho(\boldsymbol{\alpha},\boldsymbol{\gamma})$.

4. 空间的正交分解

设$\boldsymbol{W}$是酉空间$\mathbf{C}^n$的一个k维子空间，$\boldsymbol{\varepsilon}_1$，$\boldsymbol{\varepsilon}_2$，…，$\boldsymbol{\varepsilon}_k$是子空间$\boldsymbol{W}$的一个正交基，补充$n-k$个正交向量$\boldsymbol{\varepsilon}_{k+1}$，$\boldsymbol{\varepsilon}_{k+2}$，…，$\boldsymbol{\varepsilon}_n$，使$\boldsymbol{\varepsilon}_1$，$\boldsymbol{\varepsilon}_2$，…，$\boldsymbol{\varepsilon}_k$，$\boldsymbol{\varepsilon}_{k+1}$，…，$\boldsymbol{\varepsilon}_n$构成$\mathbf{C}^n$的一个正交

基. 由 $\boldsymbol{\varepsilon}_{k+1}$，$\boldsymbol{\varepsilon}_{k+2}$，…，$\boldsymbol{\varepsilon}_n$ 生成的子空间记为 Span $\{\boldsymbol{\varepsilon}_{k+1}, \boldsymbol{\varepsilon}_{k+2}, \cdots, \boldsymbol{\varepsilon}_n\}$，且 dim Span $\{\boldsymbol{\varepsilon}_{k+1}, \boldsymbol{\varepsilon}_{k+2}, \cdots, \boldsymbol{\varepsilon}_n\}=n-k$.

$\forall \boldsymbol{\alpha}\in$ Span $\{\boldsymbol{\varepsilon}_{k+1}, \boldsymbol{\varepsilon}_{k+2}, \cdots, \boldsymbol{\varepsilon}_n\}$ 与 $\boldsymbol{\beta}\in \boldsymbol{W}$，有 $\boldsymbol{\alpha}\perp\boldsymbol{\beta}$；称 Span $\{\boldsymbol{\varepsilon}_{k+1}, \boldsymbol{\varepsilon}_{k+2}, \cdots, \boldsymbol{\varepsilon}_n\}$ 为 $\boldsymbol{W}$ 的正交补空间，记为 $\boldsymbol{W}^{\perp}=$ Span $\{\boldsymbol{\varepsilon}_{k+1}, \boldsymbol{\varepsilon}_{k+2}, \cdots, \boldsymbol{\varepsilon}_n\}$，所以 $\mathbf{C}^n=\boldsymbol{W}\oplus\boldsymbol{W}^{\perp}$.

当然 $\mathbf{C}^n$ 也可以分解为若干个互相正交的子空间之直和. 特别的，对 $\mathbf{C}^n$ 中给定的正交基 $\boldsymbol{\varepsilon}_1$，$\boldsymbol{\varepsilon}_2$，…，$\boldsymbol{\varepsilon}_n$

$$\mathbf{C}^n=\text{Span}\{\boldsymbol{\varepsilon}_1\}\oplus\text{Span}\{\boldsymbol{\varepsilon}_2\}\oplus\cdots\oplus\text{Span}\{\boldsymbol{\varepsilon}_n\}$$

同理，欧氏空间也可以分解.

【例 1.5.10】 设欧氏空间 $\mathbf{R}^4$ 中标准正交基 $\boldsymbol{e}_1$，$\boldsymbol{e}_2$，$\boldsymbol{e}_3$，$\boldsymbol{e}_4$，$\mathbf{R}^4$ 的子空间 $\boldsymbol{W}=$ Span $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\}$，$\boldsymbol{\alpha}_1=\boldsymbol{e}_1+\boldsymbol{e}_2$，$\boldsymbol{\alpha}_2=\boldsymbol{e}_1+\boldsymbol{e}_2-\boldsymbol{e}_3$，试求 $\boldsymbol{W}$ 的正交补空间 $\boldsymbol{W}^{\perp}$.

解 设 $\boldsymbol{\beta}=x_1\boldsymbol{e}_1+x_2\boldsymbol{e}_2+x_3\boldsymbol{e}_3+x_4\boldsymbol{e}_4\in\boldsymbol{W}^{\perp}$，

$\because(\boldsymbol{\alpha}_i, \boldsymbol{\beta})=0$，$(i=1, 2)$

$$\therefore \left.\begin{array}{r}x_1+x_2=0\\ x_1+x_2-x_3=0\end{array}\right\}\text{，因此基础解系为}\left.\begin{array}{l}\boldsymbol{\beta}_1=(1, -1, 0, 0)^{\mathrm{T}}\\ \boldsymbol{\beta}_2=(0, 0, 0, 1)^{\mathrm{T}}\end{array}\right\}.$$

$\therefore\boldsymbol{W}^{\perp}=\text{Span}\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2\}$，$\therefore\mathbf{R}^4=\boldsymbol{W}\oplus\boldsymbol{W}^{\perp}$.

【例 1.5.11】 设 $A\in\mathbf{R}^{m\times n}$，证明：$\mathbf{R}^{\perp}(\boldsymbol{A})=\mathbf{N}(\boldsymbol{A}^{\mathrm{T}})$.

证 将矩阵 $\boldsymbol{A}$ 按列分块为 $\boldsymbol{A}=(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n)$，则 $x\in\mathbf{R}^{\perp}(\boldsymbol{A})\Leftrightarrow x\perp\mathbf{R}(\boldsymbol{A})=\text{Span}(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n)\Leftrightarrow x\perp\boldsymbol{a}_j$，$(j=1, 2, \cdots, n)\Leftrightarrow\boldsymbol{a}_j^{\mathrm{T}}x=0$，$(j=1, 2, \cdots, n)\Leftrightarrow\boldsymbol{A}^{\mathrm{T}}x=0\Leftrightarrow x\in\mathbf{N}(\boldsymbol{A}^{\mathrm{T}})$.

5. 线性空间的同构

我们已经知道，如果在数域 P 上的 n 维线性空间 $\mathbf{V}$ 中取定一个基后，$\mathbf{V}$ 中的每一个向量 $\boldsymbol{\alpha}$ 有唯一确定的坐标 $(x_1, x_2, \cdots, x_n)$ 与之对应，而向量的坐标就是数域 P 上的 n 元数组，即 $(x_1, x_2, \cdots, x_n)\in P^n$. 所以，存在着一个由 $\mathbf{V}$ 到 P^n 的映射 f. 并且 $f:\boldsymbol{\alpha}\mapsto(x_1,x_2,\cdots,x_n)$.

反之，对任意的 $(x_1,x_2,\cdots,x_n)\in P^n$，$\boldsymbol{\alpha}=\sum\limits_{i=1}^{n}x_i\boldsymbol{\alpha}_i$ 是 $\mathbf{V}$ 中唯一确定的向量，并且 $f(\boldsymbol{\alpha})=(x_1,x_2,\cdots,x_n)$. 所以映射 f 是一个由 $\mathbf{V}$ 到 P^n 上的 1－1 映射.

由前述定理可知，对任意的 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\mathbf{V}$，若 $f(\boldsymbol{\alpha})=(x_1, x_2, \cdots, x_n)$，$f(\boldsymbol{\beta})=(y_1, y_2, \cdots, y_n)$，则有 $f(\boldsymbol{\alpha}+\boldsymbol{\beta})=(x_1+y_1, x_2+y_2, \cdots, x_n+y_n)=f(\boldsymbol{\alpha})+f(\boldsymbol{\beta})$；以及对任意的 $a\in P$，有 $f(a\boldsymbol{\alpha})=(ax_1, ax_2, \cdots, ax_n)=af(\boldsymbol{\alpha})$. 也就是说映射 f 保持向量加法和纯量乘法的运算.

定义 1.5.8 设 $\mathbf{V}$，$\boldsymbol{W}$ 是数域 P 上的两个线性空间，由 $\mathbf{V}$ 到 $\boldsymbol{W}$ 的一个映射 f 叫做一个同构映射，如果：

(1) f 是 $\mathbf{V}$ 到 $\boldsymbol{W}$ 的一个 1－1 映射；

(2) $\forall\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\mathbf{V}$，$f(\boldsymbol{\alpha}+\boldsymbol{\beta})=f(\boldsymbol{\alpha})+f(\boldsymbol{\beta})$；

(3) $\forall a\in P$，$\boldsymbol{\alpha}\in\mathbf{V}$，$f(a\boldsymbol{\alpha})=af(\boldsymbol{\alpha})$.

定义 1.5.9 如果数域 P 上的两个线性空间 $\mathbf{V}$，$\boldsymbol{W}$ 之间可以建立一个同构映射，则称线性空间 $\boldsymbol{W}$ 与线性空间 $\mathbf{V}$ 同构，记作 $\mathbf{V}\cong\boldsymbol{W}$.

线性空间实质上是一个带有两个特殊运算的集合，至于集合中的元素是什么，并不重

要. 因此，从这个意义上看，同构的线性空间是一致的.

定理 1.5.6 数域 P 上任意一个 n 维线性空间都与线性空间 P^n 同构.

定理 1.5.7 设 $\boldsymbol{V}$，$\boldsymbol{W}$ 是数域 P 上的两个线性空间，f 是 $\boldsymbol{V}$ 到 $\boldsymbol{W}$ 上的一个同构映射，那么：

(1) $f(\boldsymbol{0})=\boldsymbol{0}$；

(2) $\forall\boldsymbol{\alpha}\in\boldsymbol{V}$，$f(-\boldsymbol{\alpha})=-f(\boldsymbol{\alpha})$；

(3) $f(a_1\boldsymbol{\alpha}_1+a_2\boldsymbol{\alpha}_2+\cdots+a_n\boldsymbol{\alpha}_n)=a_1f(\boldsymbol{\alpha}_1)+a_2f(\boldsymbol{\alpha}_2)+\cdots+a_nf(\boldsymbol{\alpha}_n)$（其中 $\forall a_i\in P$，$\boldsymbol{\alpha}_i\in\boldsymbol{V}$，$i=1, 2, \cdots, n$）；

(4) 向量组 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\in\boldsymbol{V}$ 线性相关的充分且必要条件为 $f(\boldsymbol{\alpha}_1), f(\boldsymbol{\alpha}_2), \cdots, f(\boldsymbol{\alpha}_n)\in\boldsymbol{W}$ 线性相关；

(5) $\boldsymbol{V}$ 到 $\boldsymbol{W}$ 上的同构映射 f 的逆映射 f^{-1} 是 $\boldsymbol{W}$ 到 $\boldsymbol{V}$ 上的同构映射.

证 (1) 在定义 1.5.8 的条件 (3) 中取 $a=0$，则 $f(\boldsymbol{0})=f(0\boldsymbol{\alpha})=0f(\boldsymbol{\alpha})=\boldsymbol{0}$；

(2) $f(\boldsymbol{\alpha})+f(-\boldsymbol{\alpha})=f(\boldsymbol{\alpha}+(-\boldsymbol{\alpha}))=f(\boldsymbol{0})=0$，所以

$$f(-\boldsymbol{\alpha})=-f(\boldsymbol{\alpha})$$

(3) 因为 $\forall\boldsymbol{\alpha}, \boldsymbol{\beta}\in\boldsymbol{V}$，有 $f(\boldsymbol{\alpha}+\boldsymbol{\beta})=f(\boldsymbol{\alpha})+f(\boldsymbol{\beta})$，以及 $\forall a\in P$，$\boldsymbol{\alpha}\in\boldsymbol{V}$，$f(a\boldsymbol{\alpha})=af(\boldsymbol{\alpha})$，应用数学归纳法可证明 $f\left(\sum\limits_{i=1}^{n}a_i\boldsymbol{\alpha}_i\right)=\sum\limits_{i=1}^{n}f(a_i\boldsymbol{\alpha}_i)=\sum\limits_{i=1}^{n}a_if(\boldsymbol{\alpha}_i)$；

(4) 若 $a_1\boldsymbol{\alpha}_1+a_2\boldsymbol{\alpha}_2+\cdots+a_n\boldsymbol{\alpha}_n=\boldsymbol{0}$，则 $a_1f(\boldsymbol{\alpha}_1)+a_2f(\boldsymbol{\alpha}_2)+\cdots+a_nf(\boldsymbol{\alpha}_n)=f(a_1\boldsymbol{\alpha}_1+a_2\boldsymbol{\alpha}_2+\cdots+a_n\boldsymbol{\alpha}_n)=f(\boldsymbol{0})=\boldsymbol{0}$，反之，若 $f(a_1\boldsymbol{\alpha}_1+a_2\boldsymbol{\alpha}_2+\cdots+a_n\boldsymbol{\alpha}_n)=\boldsymbol{0}$，因为 f 是单射，由 $f(\boldsymbol{0})=\boldsymbol{0}$，所以必有 $a_1\boldsymbol{\alpha}_1+a_2\boldsymbol{\alpha}_2+\cdots+a_n\boldsymbol{\alpha}_n=\boldsymbol{0}$，由此可知向量组 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 与向量组 $f(\boldsymbol{\alpha}_1), f(\boldsymbol{\alpha}_2), \cdots, f(\boldsymbol{\alpha}_n)$ 同时线性相关或线性无关；

(5) 因为 f 是 $\boldsymbol{V}$ 到 $\boldsymbol{W}$ 上的 1－1 映射，所以 f 的逆映射 f^{-1} 是 $\boldsymbol{W}$ 到 $\boldsymbol{V}$ 上的 1－1 映射，并且 ff^{-1} 是 $\boldsymbol{W}$ 到自身的恒等映射，而 $f^{-1}f$ 是 V 到自身的恒等映射.

设 $\boldsymbol{\alpha}'$，$\boldsymbol{\beta}'\in\boldsymbol{W}$，因为 f 是 $\boldsymbol{V}$ 到 $\boldsymbol{W}$ 上的同构映射，所以

$$f(f^{-1}(\boldsymbol{\alpha}'+\boldsymbol{\beta}'))=\boldsymbol{\alpha}'+\boldsymbol{\beta}'=f(f^{-1}(\boldsymbol{\alpha}'))+f(f^{-1}(\boldsymbol{\beta}'))=f(f^{-1}(\boldsymbol{\alpha}')+f^{-1}(\boldsymbol{\beta}'))$$

又因为 f 是单射，所以 $f^{-1}(\boldsymbol{\alpha}'+\boldsymbol{\beta}')=f^{-1}(\boldsymbol{\alpha}')+f^{-1}(\boldsymbol{\beta}')$；

同理 $\forall a\in P$，$\boldsymbol{\alpha}'\in\boldsymbol{V}$，$f^{-1}(a\boldsymbol{\alpha}')=af^{-1}(\boldsymbol{\alpha}')$.

定理 1.5.8 数域 P 上的两个有限维线性空间同构的充分且必要条件为它们有相同的维数.

证 设 $\boldsymbol{V}$，$\boldsymbol{W}$ 是数域 P 上的两个有限维线性空间，如果 $\dim\boldsymbol{V}=\dim\boldsymbol{W}=n$，令 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 与 $\{\boldsymbol{\alpha}'_1, \boldsymbol{\alpha}'_2, \cdots, \boldsymbol{\alpha}'_n\}$ 分别是 $\boldsymbol{V}$ 与 $\boldsymbol{W}$ 的基，对 $\boldsymbol{\alpha}=\sum\limits_{i=1}^{n}a_i\boldsymbol{\alpha}_i\in\boldsymbol{V}$，定义 $f(\boldsymbol{\alpha})=\sum\limits_{i=1}^{n}a_i\boldsymbol{\alpha}'_i\in\boldsymbol{W}$，易证 f 是 $\boldsymbol{V}$ 到 $\boldsymbol{W}$ 上的一个同构映射；反之，若 $\boldsymbol{V}$ 与 $\boldsymbol{W}$ 同构，令 f 是 $\boldsymbol{V}$ 到 $\boldsymbol{W}$ 上的一个同构映射，并设 $\dim\boldsymbol{V}=n$，$\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 是 $\boldsymbol{V}$ 的基，由定理 1.5.7 易证 $\{f(\boldsymbol{\alpha}_1), f(\boldsymbol{\alpha}_2), \cdots, f(\boldsymbol{\alpha}_n)\}$ 是 $\boldsymbol{W}$ 的一个基，$\therefore\dim\boldsymbol{W}=n$.

所以线性空间 P^n 可以作为数域 P 上的 n 维线性空间的代表.

习 题 1

1.1 下列映射中哪些是单射，哪些是满射，哪些是 1 - 1 映射：

(1) $\begin{array}{l} f:\mathbf{R}\to[-1,1] \\ x\mapsto f(x)=\cos x \end{array}$;

(2) $\begin{array}{l} f:\mathbf{R}\to\mathbf{R} \\ x\mapsto f(x)=2x-3 \end{array}$;

(3) $\begin{array}{l} f:\mathbf{N}\to\mathbf{N} \\ n\mapsto f(n)=2n \end{array}$;

(4) $\begin{array}{l} f:\mathbf{R}^{2\times2}\to\mathbf{R}^{2\times2} \\ X\mapsto f(X)=X\boldsymbol{P} \end{array}$,其中 $\boldsymbol{P}=\begin{pmatrix}1&2\\0&4\end{pmatrix}$；

(5) $\begin{array}{l} f:\mathbf{R}^{2\times2}\to\mathbf{R}^{2\times2} \\ X\mapsto f(X)=\boldsymbol{P}X \end{array}$，其中 $\boldsymbol{P}=\begin{pmatrix}1&2\\2&4\end{pmatrix}$；

(6) $\begin{array}{l} f:\boldsymbol{P}_n[x]\mapsto\boldsymbol{P}_n[x] \\ p(x)\mapsto p'(x) \end{array}$;

(7) $\begin{array}{l} f:\boldsymbol{P}_n[x]\mapsto\boldsymbol{P}_n[x] \\ p(x)\mapsto\int_0^{x^2-2}p(1)\mathrm{d}x \end{array}$.

1.2 设 $\boldsymbol{A}\in P^{n\times n}$ 是一个固定矩阵，令

$$\boldsymbol{V}=\{\boldsymbol{B}\in P^{n\times n}\mid \boldsymbol{AB}=\boldsymbol{BA}\}$$

证明：$\boldsymbol{V}$ 是数域 P 上的线性空间.

1.3 设 $\boldsymbol{V}=\{(a_1,a_2,\cdots,a_n)\mid a_i\in P,a_i\geqslant 0\}$，证明 $\boldsymbol{V}$ 对 n 维向量的加法与数乘不能构成线性空间.

1.4 验证 $2-t^2+3t^3$，$3-2t+t^2-t^3$，$2-t^2+4t^3$，$1-2t+3t^2-5t^3$ 是 $\boldsymbol{P}_3(t)$ 的基.

1.5 求 $\mathbf{R}^{2\times2}$ 中向量组 $\begin{pmatrix}0&1\\-1&-1\end{pmatrix}$，$\begin{pmatrix}0&3\\0&-1\end{pmatrix}$，$\begin{pmatrix}0&0\\1&3\end{pmatrix}$，$\begin{pmatrix}0&-1\\1&-1\end{pmatrix}$，$\begin{pmatrix}0&5\\-2&6\end{pmatrix}$ 的秩和一个极大线性无关组.

1.6 求 $\begin{pmatrix}2&4\\1&3\end{pmatrix}$ 在基 $\begin{pmatrix}1&1\\1&1\end{pmatrix}$，$\begin{pmatrix}1&0\\1&1\end{pmatrix}$，$\begin{pmatrix}1&1\\0&1\end{pmatrix}$，$\begin{pmatrix}0&1\\1&1\end{pmatrix}$ 下的坐标.

1.7 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，$\lambda$ 是数. 记 $\boldsymbol{E}_\lambda=\{\boldsymbol{x}\in\mathbf{C}^n\mid \boldsymbol{Ax}=\lambda\boldsymbol{x}\}$. 证明 $\boldsymbol{E}_\lambda$ 是 $\mathbf{C}^n$ 的子空间. 如果 λ 是 $\boldsymbol{A}$ 的特征值，则称 $\boldsymbol{E}_\lambda$ 是 $\boldsymbol{A}$ 的特征子空间.

1.8 设有 $\mathbf{R}^3$ 的两个子空间

$$\boldsymbol{V}_1=\{(x_1,x_2,x_3)\mid 2x_1+x_2-x_3=0\}$$
$$\boldsymbol{V}_2=\{(x_1,x_2,x_3)\mid x_1+x_2=0,3x_1+2x_2-x_3=0\}$$

求子空间 $\boldsymbol{V}_1+\boldsymbol{V}_2$，$\boldsymbol{V}_1\cap\boldsymbol{V}_2$ 的基，以及它们的维数.

1.9 已知 $\boldsymbol{A}_1=\begin{pmatrix}2&0\\-1&3\end{pmatrix}$，$\boldsymbol{A}_2=\begin{pmatrix}3&-2\\1&-1\end{pmatrix}$，$\boldsymbol{B}_1=\begin{pmatrix}-5&6\\-5&9\end{pmatrix}$，$B_2=\begin{pmatrix}4&-4\\3&-5\end{pmatrix}$，而

$$V_1 = \text{span}\{A_1, A_2\}, V_2 = \text{span}\{B_1, B_2\}$$

求子空间 V_1+V_2，$V_1\cap V_2$ 的基，以及它们的维数.

1.10 证明：span $\{\alpha_1, \alpha_2, \cdots, \alpha_r\}$ 是包含 α_1，α_2，…，α_r 的最小子空间，即若子空间 W 包含 α_1，α_2，…，α_r，则 $W=\text{span}\{\alpha_1, \alpha_2, \cdots, \alpha_r\}$.

1.11 设 A 为实对称矩阵，在什么条件下，满足 $X^TAX=0$ 的实 n 维向量 $X=(x_1,x_2,\cdots,x_n)$ 构成 $\mathbf{R}$ 上的线性空间，V 的维数如何计算?

1.12 已知 $\mathbf{R}^{2\times 2}$ 中的两组基分别是

$$e_1=\begin{pmatrix}1&0\\0&0\end{pmatrix}, e_2=\begin{pmatrix}0&0\\1&0\end{pmatrix}, e_3=\begin{pmatrix}0&1\\0&0\end{pmatrix}, e_4=\begin{pmatrix}0&0\\0&1\end{pmatrix}$$

$$\alpha_1=\begin{pmatrix}2&0\\-1&3\end{pmatrix}, \alpha_2=\begin{pmatrix}3&-2\\1&-1\end{pmatrix}, \alpha_3=\begin{pmatrix}2&0\\-1&4\end{pmatrix}, \alpha_4=\begin{pmatrix}1&-2\\3&-5\end{pmatrix}.$$

(1) 求由基 e_1，e_2，e_3，e_4 到基 α_1，α_2，α_3，α_4 的过渡矩阵；

(2) 分别求矩阵 $\begin{pmatrix}2&4\\1&3\end{pmatrix}$ 在基 e_1，e_2，e_3，e_4 和基 α_1，α_2，α_3，α_4 下的坐标；

(3) 设 V 是由在基 e_1，e_2，e_3，e_4 和基 α_1，α_2，α_3，α_4 下有相同坐标的二阶方阵构成的集合. 证明：V 是 $\mathbf{R}^{2\times 2}$ 的子空间，并求出 V 的基与维数.

1.13 设 V 是数域 P 上的 n 维线性空间，由 α_1，α_2，…，α_n 到基 β_1，β_2，…，β_n 的过渡矩阵是 A. 证明：V 中存在非零向量 η，η 在两组基下的坐标相同的充要条件是 $|E-A|=0$.

1.14 设

$$A=\begin{bmatrix}1&0&0\\0&1&0\\3&1&2\end{bmatrix}$$

(1) 证明：$V=\{X\in\mathbf{R}^{3\times 3} \mid AX=XA\}$ 是 $\mathbf{R}^{3\times 3}$ 的线性子空间.

(2) 求 V 的维数和一组基.

1.15 设

$$V_1=\{(x_1,x_2,\cdots,x_n) \mid x_1+x_2+\cdots+x_n=0\}$$
$$V_2=\{(x_1,x_2,\cdots,x_n) \mid x_1=x_2=\cdots=x_n\}$$

(1) 分别求 V_1，V_2 的基；

(2) 证明：$\mathbf{R}^n=V_1\oplus V_2$.

1.16 设

$$V_1=\{p(x) \mid p(x)\text{ 是奇函数}, p(x)\in P_n[x]\}$$
$$V_2=\{p(x) \mid p(x)\text{ 是偶函数}, p(x)\in P_n[x]\}$$

(1) 证明：V_1，V_2 是 $P_n[x]$ 的子空间，并求它们的基；

(2) 证明：$P_n[x]=V_1\oplus V_2$.

1.17 设 V_1，V_2 是 V 的两个非平凡子空间. 证明：存在 $\alpha\in V$，使得 $\alpha\notin V_1\cup V_2$.

1.18 设 V_1，V_2 是 V 的两个子空间，且 $\dim(V_1+V_2)=\dim(V_1\cap V_2)+1$. 证明：$V_1\subset V_2$ 或 $V_1\supset V_2$.

1.19 设 $\boldsymbol{A}$ 是 n 阶可逆实矩阵，任意将 $\boldsymbol{A}$ 分成两个子块 $\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_1\\ \boldsymbol{A}_2\end{pmatrix}$. 证明：$\mathbf{R}^n$ 是 $\boldsymbol{A}_1\boldsymbol{X}=0$ 的解空间 $\boldsymbol{V}_1$ 与 $\boldsymbol{A}_2\boldsymbol{X}=0$ 的解空间 $\boldsymbol{V}_2$ 的直和.

1.20 设 $\boldsymbol{A}\in\boldsymbol{P}^{n\times n}$，且 $\boldsymbol{A}^2=\boldsymbol{A}$. 证明：$\boldsymbol{P}^n$ 是齐次线性方程组 $\boldsymbol{AX}=0$ 与 $(\boldsymbol{A}-\boldsymbol{E})\boldsymbol{X}=0$ 的解空间的直和.

1.21 设 $\boldsymbol{\varepsilon}_1$，$\boldsymbol{\varepsilon}_2$，$\boldsymbol{\varepsilon}_3$ 是 3 维欧氏空间 $\boldsymbol{V}$ 的一组标准正交基，$\boldsymbol{\alpha}=3\boldsymbol{\varepsilon}_1+2\boldsymbol{\varepsilon}_2+4\boldsymbol{\varepsilon}_3$，$\boldsymbol{\beta}=\boldsymbol{\varepsilon}_1-2\boldsymbol{\varepsilon}_2$.

(1) 求与 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$ 都正交的全部元素；

(2) 求与 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$ 都正交的全部单位元素.

第二章 线 性 变 换

变换的概念是数学的一个基本概念，在解析几何与微积分学中曾多次学习和使用不同的变换. 现在我们主要研究线性空间到自身的变换，即所谓线性变换. 它实质上就是线性空间到自身的一个特殊的映射，因此线性变换表示了线性空间中的元素之间的关系. 如果所研究的线性空间是有限维线性空间的话，使用矩阵来研究线性变换是十分方便的，本章中我们将使用矩阵这一工具来研究线性变换. 如无特殊说明，以下仅讨论有限维线性空间.

§2.1 线性变换的定义

定义 2.1.1 设 f 是数域 P 上的线性空间 $\boldsymbol{V}$ 到自身的一个映射，如果下列条件被满足，则称 f 是线性空间 $\boldsymbol{V}$ 的一个线性变换.

(1) 对于任意的 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\boldsymbol{V}$，$f(\boldsymbol{\alpha}+\boldsymbol{\beta})=f(\boldsymbol{\alpha})+f(\boldsymbol{\beta})$；

(2) 对于任意的 $a\in P$，$\boldsymbol{\alpha}\in\boldsymbol{V}$，$f(a\boldsymbol{\alpha})=af(\boldsymbol{\alpha})$.

【例 2.1.1】 平面直角坐标系绕坐标原点旋转 θ 角的变换就是欧氏空间 $\mathbf{R}^2$ 的一个线性变换，对于 $\mathbf{R}^2$ 中的任意向量 $\boldsymbol{\alpha}=(x_1,\ x_2)^{\mathrm{T}}$，则这个线性变换的象就是

$$f(\boldsymbol{\alpha})=\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}$$

【例 2.1.2】 定义在闭区间$[a,\ b]$上的所有连续实函数的集合$\boldsymbol{C}[a,\ b]$是实数域 $\mathbf{R}$ 上一个线性空间，在$\boldsymbol{C}[a,\ b]$上定义变换 J：$J(f(t))=\int_0^t f(x)\mathrm{d}x,\forall f(t)\in \boldsymbol{C}[a,b]$，则 J 是$\boldsymbol{C}[a,\ b]$的一个线性变换.

以上两例可由定义 2.1.1 验证.

【例 2.1.3】 判定矩阵空间的下列变换是否为线性变换：

(1) 在 $\boldsymbol{K}^{n\times n}$中，$\boldsymbol{T}(\boldsymbol{X})=\boldsymbol{AX}-\boldsymbol{XB},\forall \boldsymbol{X}\in\boldsymbol{K}^{n\times n}$，而 $\boldsymbol{A}$，$\boldsymbol{B}\in K^{n\times n}$取定；

(2) 在 $\boldsymbol{K}^{n\times n}$中，$\boldsymbol{T}(\boldsymbol{X})=\boldsymbol{AXB}+\boldsymbol{C}$，$\forall \boldsymbol{X}\in\boldsymbol{K}^{n\times n}$，而 $\boldsymbol{A}$，$\boldsymbol{B}$，$\boldsymbol{C}$ 分别为取定的 m 阶方阵，n 阶方阵和 $m\times n$ 矩阵.

解 (1) 是. $\forall \boldsymbol{X}$，$\boldsymbol{Y}\in\boldsymbol{K}^{n\times n}$和 $k\in\boldsymbol{K}$，有：

$$\boldsymbol{T}(\boldsymbol{X}+\boldsymbol{Y})=\boldsymbol{A}(\boldsymbol{X}+\boldsymbol{Y})-(\boldsymbol{X}+\boldsymbol{Y})\boldsymbol{B}=(\boldsymbol{AX}-\boldsymbol{XB})+(\boldsymbol{AY}-\boldsymbol{YB})=\boldsymbol{T}(\boldsymbol{X}+\boldsymbol{T}(\boldsymbol{Y}))$$

$$\boldsymbol{T}(k\boldsymbol{X})=\boldsymbol{A}(k\boldsymbol{X})-(k\boldsymbol{X})\boldsymbol{B}=k(\boldsymbol{AX}-\boldsymbol{XB})=k\boldsymbol{T}(\boldsymbol{X})$$

(2) 当 $\boldsymbol{C}\neq 0$ 时，$\boldsymbol{T}$ 不是线性变换，因为 $\boldsymbol{T}(\mathbf{0})=\boldsymbol{C}\neq\mathbf{0}$；当 $\boldsymbol{C}=0$时，$\boldsymbol{T}$ 是线性变换，因为 $\boldsymbol{X}$，$\boldsymbol{Y}\in\boldsymbol{K}^{n\times n}$和 $k\in K$ 有：

$$\boldsymbol{T}(\boldsymbol{X}+\boldsymbol{Y})=\boldsymbol{A}(\boldsymbol{X}+\boldsymbol{Y})\boldsymbol{B}=\boldsymbol{A}\times\boldsymbol{B}+\boldsymbol{A}\times\boldsymbol{E}=\boldsymbol{T}(\boldsymbol{X})+\boldsymbol{T}(\boldsymbol{Y}),$$

$$\boldsymbol{T}(k\boldsymbol{X})=\boldsymbol{A}(k\boldsymbol{X})\boldsymbol{B}=k(\boldsymbol{A}\times\boldsymbol{B})=k\boldsymbol{T}(\boldsymbol{X}).$$

【例 2.1.4】 密码法是信息编码与解码的技巧，其中的一种是使用线性变换的方法. 先在不同字母与数字间建立一一对应的关系，例如

$$\begin{array}{cccccccc} a & b & c & \cdots & x & y & z \\ \updownarrow & \updownarrow & \updownarrow & & \updownarrow & \updownarrow & \updownarrow \\ 1 & 2 & 3 & \cdots & 24 & 25 & 26 \end{array}$$

设要发出信息“action”，使用上述代码，则信息的编码为：“1，3，20，9，15，14”. 它可写成向量 $\boldsymbol{\alpha}_1=(1,3,20)^{\mathrm{T}}$，$\boldsymbol{\alpha}_2=(9,15,14)^{\mathrm{T}}$. 现选取一可逆矩阵 $\boldsymbol{A}=\begin{pmatrix}1&2&3\\1&1&2\\0&1&2\end{pmatrix}$，并定义线性变换为

$$T(\boldsymbol{\alpha})=\boldsymbol{A\alpha}$$

则

$$T(\boldsymbol{\alpha}_1)=(67,44,43)^{\mathrm{T}},T(\boldsymbol{\alpha}_2)=(81,52,43)^{\mathrm{T}}$$

在此变换下明码 1，3，20，9，15，14 就变成了密码 67，44，43，81，52，43. 收到此信息后利用逆变换进行解码，就可得到原信息.

f 是数域 P 上的线性空间 $\boldsymbol{V}$ 的一个线性变换，则 f 有如下的简单性质：

(1) $\boldsymbol{f(0)=0}$，$\boldsymbol{f(-\alpha)=-f(\alpha)}$；

(2) 若 $\boldsymbol{\beta}=k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_r\boldsymbol{\alpha}_r$，则 $f(\boldsymbol{\beta})=k_1f(\boldsymbol{\alpha}_1)+k_2f(\boldsymbol{\alpha}_2)+\cdots+k_rf(\boldsymbol{\alpha}_r)$；

(3) 若 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_r$ 线性相关，则 $f(\boldsymbol{\alpha}_1)$，$f(\boldsymbol{\alpha}_2)$，…，$f(\boldsymbol{\alpha}_r)$ 也线性相关，但反之不然.

其中 $\forall\boldsymbol{\alpha}_i$，$\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\boldsymbol{V}$，$k_i\in P$，$i=1,2,\cdots,r$.

定义 2.1.2 如果对于任意的 $\boldsymbol{\alpha}\in\boldsymbol{V}$，恒有 $f(\boldsymbol{\alpha})=\boldsymbol{0}$，则 f 称为零变换，记作 O；如果对于任意的 $\boldsymbol{\alpha}\in\boldsymbol{V}$，恒有 $f(\boldsymbol{\alpha})=\boldsymbol{\alpha}$，则 f 称为恒等变换，记作 $\boldsymbol{E}$.

定义 2.1.3 设 f 是数域 P 上的线性空间 $\boldsymbol{V}$ 的一个线性变换，$\boldsymbol{V}$ 中所有向量在 f 下的象 $f(\boldsymbol{\alpha})$ 的集合称为 f 的值域，记作 $\boldsymbol{R}(f)$，即 $\boldsymbol{R}(f)=\{f(\boldsymbol{\alpha})\mid\boldsymbol{\alpha}\in\boldsymbol{V}\}\subseteq\boldsymbol{V}$，$\boldsymbol{R}(f)$ 也被称作 f 的象子空间. 在线性空间 $\boldsymbol{V}$ 中，所有被 f 变为零向量的元素所构成的集合称为 f 的核，记作 $\mathrm{Ker}(f)$ 或 $f^{-1}(0)$，$\mathrm{Ker}(f)$ 也被称作 f 的零空间或核子空间，记作 $\boldsymbol{N}(f)$，即

$$\boldsymbol{N}(f)=\mathrm{Ker}(f)=\{\boldsymbol{\alpha}\mid f(\boldsymbol{\alpha})=\boldsymbol{0},\boldsymbol{\alpha}\in\boldsymbol{V}\}\subseteq\boldsymbol{V}.$$

定理 2.1.1 线性空间 $\boldsymbol{V}$ 的线性变换 f 的值域与核都是 $\boldsymbol{V}$ 的线性子空间.

证 因为 $\boldsymbol{V}$ 非空，所以 $\boldsymbol{R}(f)$ 也非空. 又因为 $f(\boldsymbol{\alpha})+f(\boldsymbol{\beta})=f(\boldsymbol{\alpha}+\boldsymbol{\beta})$，$k(f(\boldsymbol{\alpha})=f(k\boldsymbol{\alpha})$，即 $\boldsymbol{R}(f)$ 对于线性运算封闭，所以 $\boldsymbol{R}(f)$ 是 $\boldsymbol{V}$ 的线性子空间；再由 $f(\boldsymbol{\alpha})=0$，$f(\boldsymbol{\beta})=\boldsymbol{0}$ 可知 $f(\boldsymbol{\alpha}+\boldsymbol{\beta})=\boldsymbol{0}$，$f(k\boldsymbol{\alpha})=\boldsymbol{0}$，即 $\boldsymbol{N}(f)$ 对于线性运算封闭，因为 $f(\boldsymbol{0})=\boldsymbol{0}$，所以 $\boldsymbol{0}\in\boldsymbol{N}(f)$，即 $\boldsymbol{N}(f)$ 非空，所以 $\boldsymbol{N}(f)$ 是 $\boldsymbol{V}$ 的线性子空间.

定义 2.1.4 象子空间的维数 $\dim\boldsymbol{R}(f)$ 叫做 f 的秩，记为 $\boldsymbol{r}(f)$；核子空间的维数 $\dim\boldsymbol{N}(f)$ 叫做 f 的亏度或零度，记为 $\mathrm{null}(f)$.

定理 2.1.2 设 f 是线性空间 $\boldsymbol{V}$ 上的一个线性变换，$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 是 $\boldsymbol{V}$ 的一个基，f 在此基下的矩阵为 $\boldsymbol{A}$，则有

(1) $\boldsymbol{R}(f)=\mathrm{span}\{f(\boldsymbol{\alpha}_1),f(\boldsymbol{\alpha}_2),\cdots,f(\boldsymbol{\alpha}_n)\}$；

(2) $\boldsymbol{r}(f)=\boldsymbol{r}(\boldsymbol{A})$.

证 (1) 设 $\boldsymbol{\alpha}\in\boldsymbol{V}$，并且 $\boldsymbol{\alpha}=k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_n\boldsymbol{\alpha}_n$，

则

$$f(\boldsymbol{\alpha})=k_1f(\boldsymbol{\alpha}_1)+k_2f(\boldsymbol{\alpha}_2)+\cdots+k_nf(\boldsymbol{\alpha}_n)\tag{2-1}$$

所以 $f(\boldsymbol{\alpha})\in\text{span}\{f(\boldsymbol{\alpha}_1),f(\boldsymbol{\alpha}_2),\cdots,f(\boldsymbol{\alpha}_n)\}$，

$\therefore \boldsymbol{R}(f)\subset\text{span}\{f(\boldsymbol{\alpha}_1),f(\boldsymbol{\alpha}_2),\cdots,f(\boldsymbol{\alpha}_n)\}$，

而由（1）知基象的线性组合仍是一个象，

$\therefore \boldsymbol{R}(f)\supset\text{Span}\{f(\boldsymbol{\alpha}_1),f(\boldsymbol{\alpha}_2),\cdots,f(\boldsymbol{\alpha}_n)\}$，

所以 $\boldsymbol{R}(f)=\text{Span}\{f(\boldsymbol{\alpha}_1),f(\boldsymbol{\alpha}_2),\cdots,f(\boldsymbol{\alpha}_n)\}$；

（2）由（1）可知 f 的秩等于基象组 $\{f(\boldsymbol{\alpha}_1),f(\boldsymbol{\alpha}_2),\cdots,f(\boldsymbol{\alpha}_n)\}$ 的秩，另一方面，矩阵 $\boldsymbol{A}$ 是由基象组坐标按列排列而成的，因为 $\boldsymbol{V}$ 与 P^n 同构，所以 $\boldsymbol{r}(f)=\boldsymbol{r}(\boldsymbol{A})$.

定理 2.1.3 设 f 是线性空间 $\boldsymbol{V}$ 上的一个线性变换，则 $\boldsymbol{r}(f)+\text{null}(f)=\dim\boldsymbol{V}$，亦即 $\dim\boldsymbol{R}(f)+\dim\boldsymbol{N}(f)=n$.

证 设 $\dim\boldsymbol{N}(f)=r$，在 $N(f)$ 中取一个基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_r$，将其扩充为 $\boldsymbol{V}$ 的一个基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_r$，$\boldsymbol{\alpha}_{r+1}$，…，$\boldsymbol{\alpha}_n$. 由定理 2.1.2，$\boldsymbol{R}(f)=\text{span}\{f(\boldsymbol{\alpha}_1),f(\boldsymbol{\alpha}_2),\cdots,f(\boldsymbol{\alpha}_n)\}$，但是当 $i=1,2,\cdots,r$ 时，$f(\boldsymbol{\alpha}_i)=\boldsymbol{0}$，所以 $\boldsymbol{R}(f)=\text{span}\{f(\boldsymbol{\alpha}_{r+1}),f(\boldsymbol{\alpha}_{r+2}),\cdots,f(\boldsymbol{\alpha}_n)\}$；以下证明 $\{f(\boldsymbol{\alpha}_{r+1}),f(\boldsymbol{\alpha}_{r+2}),\cdots,f(\boldsymbol{\alpha}_n)\}$ 是 $\boldsymbol{R}(f)$ 的一个基，

设

$$k_{r+1}f(\boldsymbol{\alpha}_{r+1})+k_{r+2}f(\boldsymbol{\alpha}_{r+2})+\cdots+k_nf(\boldsymbol{\alpha}_n)=\boldsymbol{0}$$

则

$$f(k_{r+1}\boldsymbol{\alpha}_{r+1}+k_{r+2}\boldsymbol{\alpha}_{r+2}+\cdots+k_n\boldsymbol{\alpha}_n)=\boldsymbol{0}$$

所以

$$k_{r+1}\boldsymbol{\alpha}_{r+1}+k_{r+2}\boldsymbol{\alpha}_{r+2}+\cdots+k_n\boldsymbol{\alpha}_n\in\boldsymbol{N}(f)$$

$$k_{r+1}\boldsymbol{\alpha}_{r+1}+k_{r+2}\boldsymbol{\alpha}_{r+2}+\cdots+k_n\boldsymbol{\alpha}_n=k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_r\boldsymbol{\alpha}_r$$

又因为 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 线性无关，所以 $k_i=0$（$i=1,2,\cdots,n$），由此可知 $f(\boldsymbol{\alpha}_{r+1})$，$f(\boldsymbol{\alpha}_{r+2}),\cdots,f(\boldsymbol{\alpha}_n)$ 也线性无关，所以 $\boldsymbol{r}(f)=n-r$，

由此可得

$$\boldsymbol{r}(f)+\text{null}(f)=\dim\boldsymbol{V}$$

事实 虽然 $\dim\boldsymbol{R}(f)+\dim\boldsymbol{N}(f)=\dim\boldsymbol{V}=n$，但是 $\boldsymbol{R}(f)+\boldsymbol{N}(f)$ 并不一定等于 $\boldsymbol{V}$.

例如，在 $P_n[x]$ 中，令 $f(p)=p'$，则 $\boldsymbol{R}(f)=\boldsymbol{P}_{n-1}[x]$，$\boldsymbol{N}(f)=\boldsymbol{R}$，显然 $\boldsymbol{R}(f)+\boldsymbol{N}(f)\neq\boldsymbol{P}_n[x]$.

【例 2.1.5】 已知 $\mathbf{R}^{2\times2}$ 的线性变换 $\boldsymbol{T}(\boldsymbol{X})=\boldsymbol{M}\boldsymbol{X}-\boldsymbol{X}\boldsymbol{M}$，$(\forall\boldsymbol{X}\in\mathbf{R}^{2\times2},\boldsymbol{M}=\begin{pmatrix}1&2\\0&3\end{pmatrix})$，求 $\boldsymbol{R}(\boldsymbol{T})$ 与 $\boldsymbol{N}(\boldsymbol{T})$ 的基与维数.

解 取 $\mathbf{R}^{2\times2}$ 的基 E_{11}，E_{12}，E_{21}，E_{22}，则

$$\boldsymbol{T}(E_{11})=\begin{pmatrix}0&-2\\0&0\end{pmatrix},\boldsymbol{T}(E_{12})=\begin{pmatrix}0&-2\\0&0\end{pmatrix},\boldsymbol{T}(E_{21})=\begin{pmatrix}2&0\\2&-2\end{pmatrix},\boldsymbol{T}(E_{22})=\begin{pmatrix}0&2\\0&0\end{pmatrix}$$

观察可知该矩阵组的秩是 2，并且 $\begin{pmatrix}0&-2\\0&0\end{pmatrix}$，$\begin{pmatrix}2&0\\2&-2\end{pmatrix}$ 是一个极大无关组，所以 $\dim\boldsymbol{R}(T)=2$. 并且 $\boldsymbol{T}(E_{11})=\begin{pmatrix}0&-2\\0&0\end{pmatrix}$，$\boldsymbol{T}(E_{21})=\begin{pmatrix}2&0\\2&-2\end{pmatrix}$ 是 $\boldsymbol{R}(\boldsymbol{T})$ 的一个基.

设 $X=\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \in N\ (T)$，则由

$$T(X)=MX-XM=\begin{bmatrix} 2x_3 & -2x_1 & -2x_2 & +2x_4 \\ 2x_3 & & -2x_3 & \end{bmatrix}=\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

得到

$$\begin{cases} 2x_3=0 \\ -2x_1-2x_2+2x_4=0 \\ 2x_3=0 \\ -2x_3=0 \end{cases}$$

可求得该方程组的通解为

$$x_1=-k_1+k_2, x_2=k_1, x_3=0, x_4=k_2 (k_1, k_2 \text{ 任意})$$

所以

$$X=\begin{bmatrix} -k_1+k_2 & k_1 \\ 0 & k_2 \end{bmatrix} (k_1, k_2 \text{ 任意})$$

即 $\dim N(T)=2$，并且 $\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$，$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ 是 $N(T)$ 的一个基.

定义 2.1.5 设 f 是线性空间 V 上的一个线性变换，W 是 V 的一个子空间，如果对任意的向量 $\alpha \in W$，有 $f(\alpha) \in W$，则称 W 是 f 的不变子空间，简称 f-子空间.

【例 2.1.6】 线性空间 V 自身和零空间 $\{0\}$ 在任何线性变换 f 之下显然不变.

【例 2.1.7】 设 f 是线性空间 V 的一个线性变换，则 $R(f)$，$N(f)$ 是 f 的不变子空间. 这是因为 $\forall \alpha \in N(f)$，都有 $f(\alpha)=0 \in N(f)$. $R(f)=\{f\alpha \mid \alpha \in V\}$ 显然包含了 $R(f)$ 中向量的象.

【例 2.1.8】 设 3 维线性空间 V 的线性变换 T 在基 α_1，α_2，α_3 下的矩阵为

$$A=\begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$$

试证明：$W=\text{span}(-\alpha_1+\alpha_2, -\alpha_1+\alpha_3)$ 是 T 的不变子空间.

证 由 $T(\alpha_1, \alpha_2, \alpha_3)=(\alpha_1, \alpha_2, \alpha_3)A$ 得到

$$T(\alpha_1)=\alpha_1+2\alpha_2+2\alpha_3, T(\alpha_2)=2\alpha_1+\alpha_2+2\alpha_3, T(\alpha_3)=2\alpha_1+2\alpha_2+\alpha_3$$

任取 $\beta \in W$，则有

$$\beta=k_1(-\alpha_1+\alpha_2)+k_2(-\alpha_1+\alpha_3)$$

而

$$\begin{aligned} T(\beta) &= k_1[-T(\alpha_1)+T(\alpha_2)]+k_2[-T(\alpha_1)+T(\alpha_3)]=k_1(\alpha_1-\alpha_2)+k_2(\alpha_1-\alpha_3) \\ &= -k_1(-\alpha_1+\alpha_2)-k_2-(-\alpha_1+\alpha_3) \in W \end{aligned}$$

所以 W 是 T 的不变子空间.

【例 2.1.9】 设 f，g 是线性空间 V 的两个线性变换，并且 $fg=gf$，则 $R(g)$，$N(g)$ 均为 f-子空间.

$\forall \alpha \in N(g)$，有 $g[f(\alpha)]=fg(\alpha)=f[g(\alpha)]=f(0)=0$，$\therefore f(\alpha) \in N(g)$，则 $N(g)$ 为 f-子空间；$\forall g(\beta) \in R(g)$，有 $f[g(\beta)]=g[f(\beta)] \in R(g)$，$\therefore f(\beta) \in R(g)$，所以

$\boldsymbol{R}(g)$ 为 f-子空间.

【例 2.1.10】 如果 $\boldsymbol{V}$ 的子空间 $W = \text{span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots \boldsymbol{\alpha}_r\}$，则 $\boldsymbol{W}$ 是 f-子空间的充分且必要条件是 $f(\boldsymbol{\alpha}_1), f(\boldsymbol{\alpha}_2), \cdots, f(\boldsymbol{\alpha}_r) \in \boldsymbol{W}$.

证 （必要性）显然成立；

（充分性）若 $f(\boldsymbol{\alpha}_i) \in \boldsymbol{W}$，$\forall \boldsymbol{\alpha} \in \boldsymbol{W}$，因为 $\boldsymbol{\alpha} = \sum_{i=1}^{r} k_i \boldsymbol{\alpha}_i$，

所以 $f(\boldsymbol{\alpha}) = \sum_{i=1}^{r} k_i f(\boldsymbol{\alpha}_i) \in \boldsymbol{W}, i = 1, 2, \cdots, r$.

§2.2 线性变换的运算

设 $\boldsymbol{L}(\boldsymbol{V})$ 表示线性空间 $\boldsymbol{V}$ 的一切线性变换所组成的集合. 对任意的 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \boldsymbol{V}$，$k \in P$，$f$，$g \in \boldsymbol{L}(\boldsymbol{V})$，有以下的定义.

定义 2.2.1 （1）如果 $f(\boldsymbol{\alpha}) = g(\boldsymbol{\alpha})$，则称 f 与 g 相等，记作 $f = g$；

（2）$f(\boldsymbol{\alpha}) + g(\boldsymbol{\alpha})$ 称为 f 与 g 的和，记作 $f+g$，即 $(f+g)(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha}) + g(\boldsymbol{\alpha})$；

（3）$kf(\boldsymbol{\alpha})$ 称为 k 与 f 的数乘法，记作 kf，即 $(kf)(\boldsymbol{\alpha}) = kf(\boldsymbol{\alpha})$；

（4）$f(g(\boldsymbol{\alpha}))$ 称为 f 与 g 的积，记作 fg，即 $(fg)(\boldsymbol{\alpha}) = f(g(\boldsymbol{\alpha}))$.

定理 2.2.1 $\boldsymbol{L}(\boldsymbol{V})$ 对于加法与数乘法作成数域 P 上的一个线性空间.

证 设 $\forall \boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \boldsymbol{V}$，$\forall a$，$b \in \boldsymbol{R}$，$\forall f$，$g \in \boldsymbol{L}(\boldsymbol{V})$，

令 $h = f + g$，则

$$\begin{aligned} h(a\boldsymbol{\alpha} + b\boldsymbol{\beta}) &= f(a\boldsymbol{\alpha} + b\boldsymbol{\beta}) + g(a\boldsymbol{\alpha} + b\boldsymbol{\beta}) = af(\boldsymbol{\alpha}) + bf(\boldsymbol{\beta}) + ag(\boldsymbol{\alpha}) + bg(\boldsymbol{\beta}) \\ &= a(f(\boldsymbol{\alpha}) + g(\boldsymbol{\alpha})) + b(f(\boldsymbol{\beta}) + g(\boldsymbol{\beta})) = ah(\boldsymbol{\alpha}) + bh(\boldsymbol{\beta}) \end{aligned}$$

所以 $h = f + g$ 是 $\boldsymbol{V}$ 的一个线性变换，线性变换的加法满足交换律和结合律，易证以下等式成立：

（1）$f + g = g + f$；

（2）$(f + g) + l = f + (g + l)$；

（3）以 $\boldsymbol{0}$ 表示 $\boldsymbol{V}$ 的零变换，显然有

$$\boldsymbol{0} + f = f$$

（4）f 的负变换 $-f$ 是指：$-f$：$\boldsymbol{\alpha} \mapsto -f(\boldsymbol{\alpha})$，显然有

$$f + (-f) = \boldsymbol{0}$$

由此可以定义线性变换的差：$f - g = f + (-g)$，因此在 $\boldsymbol{L}(\boldsymbol{V})$ 中，加法的逆运算减法可以施行.

再设 $\forall k$，$t \in \boldsymbol{R}$，$\forall f \in \boldsymbol{L}(\boldsymbol{V})$，$\forall \boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \boldsymbol{V}$，$\forall a$，$b \in \boldsymbol{R}$.

令 $h = kf$，则

$$\begin{aligned} h(a\boldsymbol{\alpha} + b\boldsymbol{\beta}) &= k(f(a\boldsymbol{\alpha} + b\boldsymbol{\beta})) = k(af(\boldsymbol{\alpha}) + bf(\boldsymbol{\beta})) \\ &= akf(\boldsymbol{\alpha}) + bkf(\boldsymbol{\beta}) = ah(\boldsymbol{\alpha}) + bh(\boldsymbol{\beta}) \end{aligned}$$

易证下列等式成立：

（5）$k(f + g) = kf + kg$；

（6）$(k + t)f = kf + tf$；

(7) $(kt)f=k(tf)$;

(8) $1\cdot f=f$.

所以，由线性空间的定义可知$\boldsymbol{L}(\boldsymbol{V})$对于加法与数乘法作成数域$P$上的一个线性空间.

对于线性变换积的运算，以下算律成立：

(9) $f(g+l)=fg+fl$;

(10) $(g+l)f=gf+lf$;

(11) $(kf)g=k(fg)$;

(12) $(fg)l=f(gl)$.

算律 (1) ～ (12) 中 f，g，$l\in\boldsymbol{L}(\boldsymbol{V})$.

由算律 (12) 可以定义线性变换的幂运算，即 f 的 n 次幂 $f^n=\overbrace{f\cdot f\cdots\cdots f}^{n}$，$n\in\mathbf{Z}^+$.

令 $\boldsymbol{I}$ 表示 V 的单位变换，定义 $f^0=\boldsymbol{I}$，由此可知，$\forall f\in\boldsymbol{L}(\boldsymbol{V})$，$f$ 的任意非负整数幂有意义.

事实 $\forall f$，$g\in\boldsymbol{L}(\boldsymbol{V})$，一般的 $fg\neq gf$.

事实 若 f，$g\in\boldsymbol{L}(\boldsymbol{V})$，并且 $fg=gf=I$，则称 f 与 g 互为逆变换，记为 $f^{-1}=g$，$g^{-1}=f$.

§2.3 线性变换与矩阵

设 f 是数域 P 上 n 维线性空间 $\boldsymbol{V}$ 的一个线性变换，取 $\boldsymbol{V}$ 的一个基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$. 设 $\boldsymbol{\alpha}$ 是 $\boldsymbol{V}$ 中任意一个向量，则 $\boldsymbol{\alpha}=x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_n\boldsymbol{\alpha}_n$，这个向量在 f 的作用下有 $f(\boldsymbol{\alpha})=y_1\boldsymbol{\alpha}_1+y_2\boldsymbol{\alpha}_2+\cdots+y_n\boldsymbol{\alpha}_n$.

那么此时 $f(\boldsymbol{\alpha})$ 在基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 下的坐标 $(y_1, y_2, \cdots, y_n)^{\mathrm{T}}$ 如何计算呢？它与 $\boldsymbol{\alpha}$ 在同一个基下的坐标之间又有什么关系呢？我们先来看基向量 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 在 f 下的表示，不妨设

$$\begin{cases}f(\boldsymbol{\alpha}_1)=a_{11}\boldsymbol{\alpha}_1+a_{21}\boldsymbol{\alpha}_2+\cdots+a_{n1}\boldsymbol{\alpha}_n\\ f(\boldsymbol{\alpha}_2)=a_{12}\boldsymbol{\alpha}_1+a_{22}\boldsymbol{\alpha}_2+\cdots+a_{n2}\boldsymbol{\alpha}_n\\ \quad\cdots\\ f(\boldsymbol{\alpha}_n)=a_{1n}\boldsymbol{\alpha}_1+a_{2n}\boldsymbol{\alpha}_2+\cdots+a_{m}\boldsymbol{\alpha}_n\end{cases}$$

式中：$(a_{1j}, a_{2j}, \cdots, a_{nj})^{\mathrm{T}}$ 就是 $f(\boldsymbol{\alpha}_j)$ 在基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 下的坐标，i，$j=1$，2，…，n.

令 $A=\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{m}\end{pmatrix}$.

这个矩阵称作线性变换 f 关于基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 的矩阵，其中第 j 列元素就是 $f(\boldsymbol{\alpha}_j)$ 在基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 下的坐标.

当取定数域 P 上的 n 维线性空间 $\boldsymbol{V}$ 的一个基后，对于 $\boldsymbol{V}$ 的每一个线性变换，数域 P 上都有一个唯一确定的矩阵与之对应，因此，我们在矩阵与线性变换之间建立了一一对应

关系.

【例 2.3.1】 已知 $P_3[t]$ 的线性变换

$$T(a_0+a_1t+a_2t^2+a_3t^3)=(a_0-a_2)+(a_1-a_3)t+(a_2-a_0)t^2+(a_3-a_1)t^3$$

求 T 在基 1，t，t^2，t^3 下的矩阵.

解 ∵

$$T(1)=1-t^2, T(t)=t-t^3, T(t^2)=-1+t^2, T(t^3)=-1+t^3$$

∴T 在基 1，t，t^2，t^3 下的矩阵为 $A=\begin{bmatrix}1&0&-1&0\\0&1&0&-1\\-1&0&1&0\\0&-1&0&1\end{bmatrix}$.

【例 2.3.2】 已知 $R^{2\times2}$ 的线性变换 $T(X)=MX-XM$，($\forall X\in R^{2\times2}$，$M=\begin{pmatrix}1&2\\0&3\end{pmatrix}$)，求 $R(T)$ 与 $N(T)$ 的基与维数.

解 ∵

$$T(E_{11})=ME_{11}-E_{11}M=\begin{pmatrix}0&-2\\0&0\end{pmatrix}=-2E_{12}$$

$$T(E_{12})=ME_{12}-E_{12}M=\begin{pmatrix}0&-2\\0&0\end{pmatrix}=-2E_{12}$$

$$T(E_{21})=ME_{21}-E_{21}M=\begin{pmatrix}2&0\\2&-2\end{pmatrix}=2E_{11}-2E_{21}-2E_{22}$$

$$T(E_{22})=ME_{22}-E_{22}M=\begin{pmatrix}0&2\\0&0\end{pmatrix}=2E_{12}$$

又∵T 在基 E_{11}，E_{12}，E_{21}，E_{22} 下的矩阵为

$$A=\begin{bmatrix}0&0&2&0\\-2&-2&0&2\\0&0&2&0\\0&0&-2&0\end{bmatrix}$$

可求得 $\text{rank}A=2$，且 $(0,\ -2,\ 0,\ 0)^T$，$(2,\ 0,\ 2,\ -2)^T$ 是 A 的列向量组的一个极大无关组，故 $\dim R(T)=2$，且 $T(E_{11})=-2E_{12}=\begin{pmatrix}0&-2\\0&0\end{pmatrix}$，$T(E_{21})=2E_{11}+2E_{21}-2E_{22}=\begin{pmatrix}2&0\\2&-2\end{pmatrix}$ 是 $R(T)$ 的一组基；求解 $Ax=0$ 的基础解系 $(-1,\ 1,\ 0,\ 0)^T$，$(1,\ 0,\ 0,\ 1)^T$，故 $\dim N(T)=2$，且 $-E_{11}+E_{12}=\begin{pmatrix}-1&1\\0&0\end{pmatrix}$，$E_{11}+E_{22}=\begin{pmatrix}1&0\\0&1\end{pmatrix}$ 是 $N(T)$ 的一组基.

定理 2.3.1 设 f 是 n 维线性空间 V 的一个线性变换，f 关于基 α_1，α_2，…，α_n 的矩阵为 A，并且对任意的 α，$f(\alpha)\in V$，有 $\alpha=x_1\alpha_1+x_2\alpha_2+\cdots+x_n\alpha_n$ 且 $f(\alpha)=y_1\alpha_1+y_2\alpha_2+\cdots+y_n\alpha_n$，则

$$(y_1,y_2,\cdots,y_n)^T=A(x_1,x_2,\cdots,x_n)^T$$

证 $\because f(\boldsymbol{\alpha})=f(\sum_{i=1}^{n} x_i\boldsymbol{\alpha}_i)=\sum_{i=1}^{n} x_i f(\boldsymbol{\alpha}_i)=\sum_{i=1}^{n} x_i \sum_{j=1}^{n} a_{ji}\boldsymbol{\alpha}_j=\sum_{j=1}^{n}(\sum_{i=1}^{n} a_{ji}x_i)\boldsymbol{\alpha}_j$，

$\therefore y_i=\sum_{i=1}^{n} a_{ji}x_i,(j=1,2,\cdots,n)$.

则 $(y_1,y_2,\cdots,y_n)^{\mathrm{T}}=\boldsymbol{A}(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，此式也叫做 f 的解析表示.

定理 2.3.2 设在线性空间 $\boldsymbol{V}$ 中，由基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 到基 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，…，$\boldsymbol{\beta}_n$ 的过渡矩阵为 $\boldsymbol{P}$，$\boldsymbol{V}$ 中的线性变换 f 在这两个基下的矩阵分别为 $\boldsymbol{A}$，$\boldsymbol{B}$，则 $\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{AP}$.

证 $\because(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{P}$，并且

$$f(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A}$$
$$f(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\boldsymbol{B}$$

又$\because\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，…，$\boldsymbol{\beta}_n$ 线性无关，所以 $\boldsymbol{P}$ 可逆，有

$$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\boldsymbol{P}^{-1}$$

因此

$$\begin{aligned}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\boldsymbol{B}&=f(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)=f((\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{P})=(f(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n))\boldsymbol{P}\\&=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{AP}=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\boldsymbol{P}^{-1}\boldsymbol{AP},\end{aligned}$$

即

$$\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{AP}.$$

由上述定理可知 $\boldsymbol{V}$ 中的线性变换 f 在这两个不同的基下的矩阵相似，即 $\boldsymbol{A}\backsim\boldsymbol{B}$.

定理 2.3.3 设 $\boldsymbol{A}$，$\boldsymbol{B}$ 为线性变换 f_1，f_2 在基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 下的矩阵，那么在此基下有：

(1) f_1+f_2 在基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 下的矩阵为 $\boldsymbol{A}+\boldsymbol{B}$；

(2) kf_1 在基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 下的矩阵为 $k\boldsymbol{A}$；

(3) f_1f_2 在基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 下的矩阵为 $\boldsymbol{AB}$；

(4) 若 f_1 可逆，那么 f_1^{-1}在基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 下的矩阵为 $\boldsymbol{A}^{-1}$.

【例 2.3.3】 已知 $\mathbf{R}^{2\times2}$ 的两个线性变换 $T(\boldsymbol{X})=\boldsymbol{XN},S(\boldsymbol{X})=\boldsymbol{MX},(\forall\boldsymbol{X}\in\mathbf{R}^{2\times2},\boldsymbol{M}=\begin{pmatrix}1&0\\-2&0\end{pmatrix},\boldsymbol{N}=\begin{pmatrix}1&1\\1&-1\end{pmatrix})$. 试求：

(1) $\boldsymbol{T}+\boldsymbol{S}$，$\boldsymbol{TS}$ 在基 $\boldsymbol{E}_{11}$，$\boldsymbol{E}_{12}$，$\boldsymbol{E}_{21}$，$\boldsymbol{E}_{22}$ 下的矩阵；

(2) $\boldsymbol{T}$ 与 $\boldsymbol{S}$ 是否可逆？若可逆，求其逆变换。

解 (1) 可求得 $\boldsymbol{T}$ 与 $\boldsymbol{S}$ 在基 $\boldsymbol{E}_{11}$，$\boldsymbol{E}_{12}$，$\boldsymbol{E}_{21}$，$\boldsymbol{E}_{22}$ 下的矩阵分别为

$$\boldsymbol{A}=\begin{pmatrix}1&1&0&0\\1&-1&0&0\\0&0&1&1\\0&0&1&-1\end{pmatrix},\boldsymbol{B}=\begin{pmatrix}1&0&0&0\\0&1&0&0\\-2&0&0&0\\0&-2&0&0\end{pmatrix}$$

故 $\boldsymbol{T}+\boldsymbol{S}$，$\boldsymbol{TS}$ 在该基下的矩阵分别为

$$\boldsymbol{A}+\boldsymbol{B}=\begin{pmatrix}2&1&0&0\\1&0&0&0\\-2&0&1&1\\0&-2&1&-1\end{pmatrix},\boldsymbol{AB}=\begin{pmatrix}1&1&0&0\\1&-1&0&0\\-2&-2&0&0\\-2&2&0&0\end{pmatrix}$$

(2) 由于 $\det\boldsymbol{A}=4$，$\det\boldsymbol{B}=0$，所以线性变换 $\boldsymbol{T}$ 可逆，$\boldsymbol{S}$ 不可逆求得 $\boldsymbol{T}^{-1}$在基 $\boldsymbol{E}_{11}$，$\boldsymbol{E}_{12}$，

$\boldsymbol{E}_{21}$，$\boldsymbol{E}_{22}$下的矩阵为

$$\boldsymbol{A}^{-1}=\begin{pmatrix}1/2 & 1/2 & 0 & 0\\ 1/2 & -1/2 & 0 & 0\\ 0 & 0 & 1/2 & 1/2\\ 0 & 0 & 1/2 & -1/2\end{pmatrix}$$

任取 $\boldsymbol{X}=\begin{pmatrix}x_1 & x_2\\ x_3 & x_4\end{pmatrix}\in\mathbf{R}^{2\times 2}$，则

$$\boldsymbol{T}^{-1}(\boldsymbol{X})=\boldsymbol{T}^{-1}\left[(\boldsymbol{E}_{11},\boldsymbol{E}_{12},\boldsymbol{E}_{21},\boldsymbol{E}_{22})\begin{pmatrix}x_1\\ x_2\\ x_3\\ x_4\end{pmatrix}\right]=(\boldsymbol{E}_{11},\boldsymbol{E}_{12},\boldsymbol{E}_{21},\boldsymbol{E}_{22})\boldsymbol{A}^{-1}\begin{pmatrix}x_1\\ x_2\\ x_3\\ x_4\end{pmatrix}$$

$$=\frac{1}{2}(x_1+x_2)\boldsymbol{E}_{11}+\frac{1}{2}(x_1-x_2)\boldsymbol{E}_{12}+\frac{1}{2}(x_3+x_4)\boldsymbol{E}_{21}+\frac{1}{2}(x_3-x_4)\boldsymbol{E}_{22}$$

$$=\frac{1}{2}\begin{pmatrix}x_1+x_2 & x_1-x_2\\ x_3+x_4 & x_3-x_4\end{pmatrix}=\begin{pmatrix}x_1 & x_2\\ x_3 & x_4\end{pmatrix}\frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}=\boldsymbol{X}\boldsymbol{N}^{-1}$$

【例 2.3.4】 在三维空间 $\mathbf{R}^3$ 中求下列各线性变换 f 在指定的基下的矩阵：

(1) 已知 $f(x_1, x_2, x_3)=(2x_1-x_2, x_2+x_3, x_1)$，求 f 在基 $\boldsymbol{\varepsilon}_1=(1, 0, 0)$，$\boldsymbol{\varepsilon}_2=(0, 1, 0)$，$\boldsymbol{\varepsilon}_3=(0, 0, 1)$下的矩阵；

(2) 已知线性变换 f 在基 $\boldsymbol{\eta}_1=(-1, 1, 1)$，$\boldsymbol{\eta}_2=(1, 0, -1)$，$\boldsymbol{\eta}_3=(0, 1, 1)$下的矩阵是 $\begin{pmatrix}1 & 0 & 1\\ 1 & 1 & 0\\ -1 & 2 & 1\end{pmatrix}$，求 f 在基 $\boldsymbol{\varepsilon}_1=(1, 0, 0)$，$\boldsymbol{\varepsilon}_2=(0, 1, 0)$，$\boldsymbol{\varepsilon}_3=(0, 0, 1)$下的矩阵；

(3) 已知 $f(\boldsymbol{\eta}_1)=(-5,0,3)$，$f(\boldsymbol{\eta}_2)=(0,-1,6)$，$f(\boldsymbol{\eta}_3)=(-5,-1,9)$，其中 $\boldsymbol{\eta}_1=(-1, 0, 2)$，$\boldsymbol{\eta}_2=(0, 1, 1)$，$\boldsymbol{\eta}_3=(3, -1, 0)$是一个基，求 f 在 $\boldsymbol{\varepsilon}_1=(1, 0, 0)$，$\boldsymbol{\varepsilon}_2=(0, 1, 0)$，$\boldsymbol{\varepsilon}_3=(0, 0, 1)$下的矩阵以及在基 $\boldsymbol{\eta}_1=(-1, 0, 2)$，$\boldsymbol{\eta}_2=(0, 1, 1)$，$\boldsymbol{\eta}_3=(3, -1, 0)$下的矩阵.

解 (1) $\because f(x_1, x_2, x_3)=(2x_1-x_2, x_2+x_3, x_1)$，

$\therefore f(\boldsymbol{\varepsilon}_1)=f(1, 0, 0)=(2, 0, 1)$，$f(\boldsymbol{\varepsilon}_2)=f(0, 1, 0)=(-1, 1, 0)$，$f(\boldsymbol{\varepsilon}_3)=f(0, 0, 1)=(0, 1, 0)$.

由于三维向量在标准正交基 $\boldsymbol{\varepsilon}_1$，$\boldsymbol{\varepsilon}_2$，$\boldsymbol{\varepsilon}_3$ 下的坐标就是其分量，所以

$$f(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3)=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3)\begin{pmatrix}2 & -1 & 0\\ 0 & 1 & 1\\ 1 & 0 & 0\end{pmatrix}$$

因此 f 在基 $\boldsymbol{\varepsilon}_1$，$\boldsymbol{\varepsilon}_2$，$\boldsymbol{\varepsilon}_3$ 下的矩阵为 $\begin{pmatrix}2 & -1 & 0\\ 0 & 1 & 1\\ 1 & 0 & 0\end{pmatrix}$.

(2) $\because f(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3)=(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3)\begin{pmatrix}1 & 0 & 1\\ 1 & 1 & 0\\ -1 & 2 & 1\end{pmatrix}$，

又$(\boldsymbol{\eta}_1,\ \boldsymbol{\eta}_2,\ \boldsymbol{\eta}_3)=(\boldsymbol{\varepsilon}_1,\ \boldsymbol{\varepsilon}_2,\ \boldsymbol{\varepsilon}_3)\begin{pmatrix}-1&1&0\\1&0&1\\1&-1&1\end{pmatrix}$,

$$\therefore(\boldsymbol{\varepsilon}_1,\ \boldsymbol{\varepsilon}_2,\ \boldsymbol{\varepsilon}_3)=(\boldsymbol{\eta}_1,\ \boldsymbol{\eta}_2,\ \boldsymbol{\eta}_3)\begin{pmatrix}-1&1&0\\1&0&1\\1&-1&1\end{pmatrix}^{-1}=(\boldsymbol{\eta}_1,\ \boldsymbol{\eta}_2,\ \boldsymbol{\eta}_3)\begin{pmatrix}-1&1&-1\\0&1&-1\\1&0&1\end{pmatrix}.$$

由此可得

$$f(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3)=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3)\begin{pmatrix}-1&1&0\\1&0&1\\1&-1&1\end{pmatrix}\begin{pmatrix}1&0&1\\1&1&0\\-1&2&1\end{pmatrix}\begin{pmatrix}-1&1&-1\\0&1&-1\\1&0&1\end{pmatrix}$$

$$=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3)\begin{pmatrix}-1&1&-2\\2&2&0\\3&0&2\end{pmatrix}$$

因此 f 在基$\boldsymbol{\varepsilon}_1$，$\boldsymbol{\varepsilon}_2$，$\boldsymbol{\varepsilon}_3$ 下的矩阵为$\begin{pmatrix}-1&1&-2\\2&2&0\\3&0&2\end{pmatrix}$.

(3) $\because f(\boldsymbol{\eta}_1,\ \boldsymbol{\eta}_2,\ \boldsymbol{\eta}_3)=(\boldsymbol{\varepsilon}_1,\ \boldsymbol{\varepsilon}_2,\ \boldsymbol{\varepsilon}_3)\begin{pmatrix}-5&0&-5\\0&-1&-1\\3&6&9\end{pmatrix}$,

又$\because(\boldsymbol{\eta}_1,\ \boldsymbol{\eta}_2,\ \boldsymbol{\eta}_3)=(\boldsymbol{\varepsilon}_1,\ \boldsymbol{\varepsilon}_2,\ \boldsymbol{\varepsilon}_3)\begin{pmatrix}-1&0&3\\0&1&-1\\2&1&0\end{pmatrix}$,

$$f(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3)=f(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\eta}_3)\begin{pmatrix}-1&0&3\\0&1&-1\\2&1&0\end{pmatrix}^{-1}$$

$$=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3)\begin{pmatrix}-5&0&-5\\0&-1&-1\\3&6&9\end{pmatrix}\begin{pmatrix}-1&0&3\\0&1&-1\\2&1&0\end{pmatrix}^{-1}$$

$$=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3)\begin{pmatrix}-5&0&-5\\0&-1&-1\\3&6&9\end{pmatrix}\begin{pmatrix}-1/7&-3/7&3/7\\2/7&6/7&1/7\\2/7&-1/7&1/7\end{pmatrix}$$

$$=\frac{1}{7}(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3)\begin{pmatrix}-5&20&-20\\-4&-5&-2\\27&18&24\end{pmatrix}$$

因此 f 在基$\boldsymbol{\varepsilon}_1$，$\boldsymbol{\varepsilon}_2$，$\boldsymbol{\varepsilon}_3$ 下的矩阵为

$$\frac{1}{7}\begin{pmatrix}-5&20&-20\\-4&-5&-2\\27&18&24\end{pmatrix}$$

$$\text{又}\because f(\boldsymbol{\eta}_1,\ \boldsymbol{\eta}_2,\ \boldsymbol{\eta}_3)=(\boldsymbol{\eta}_1,\ \boldsymbol{\eta}_2,\ \boldsymbol{\eta}_3)\begin{pmatrix}-1&0&3\\0&1&-1\\2&1&0\end{pmatrix}^{-1}\begin{pmatrix}-5&0&-5\\0&-1&-1\\3&6&9\end{pmatrix}$$

$$=\frac{1}{7}(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\eta}_3)\begin{pmatrix}-1&-3&3\\2&6&1\\2&-1&1\end{pmatrix}\begin{pmatrix}-5&0&-5\\0&-1&-1\\3&6&9\end{pmatrix}=(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\eta}_3)\begin{pmatrix}2&3&5\\-1&0&-1\\-1&1&0\end{pmatrix},$$

$\therefore$ f 在基 $\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\eta}_3$ 下的矩阵为

$$\begin{pmatrix}2&3&5\\-1&0&-1\\-1&1&0\end{pmatrix}$$

§2.4　正交变换与正交矩阵

我们知道对笛卡尔平面直角坐标系而言，经过旋转变换，向量的长度保持不变．具备这种性质的线性变换有着广泛的应用．本节将在欧氏空间中讨论这类线性变换．

定义 2.4.1　设 $\mathbf{V}$ 是一个欧氏空间，f 是 $\mathbf{V}$ 的一个线性变换，如果对于任意的 $\boldsymbol{\alpha}\in\mathbf{V}$，有 $|f(\boldsymbol{\alpha})|=|\boldsymbol{\alpha}|$，或用内积的形式表示为 $(f(\boldsymbol{\alpha}),f(\boldsymbol{\alpha}))=(\boldsymbol{\alpha},\boldsymbol{\alpha})$，则 f 称作 $\mathbf{V}$ 的一个正交变换．

定理 2.4.1　欧氏空间 $\mathbf{V}$ 的一个线性变换 f 是正交变换的充分且必要条件是：对于任意的向量 $\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbf{V}$，有 $(f(\boldsymbol{\alpha}),f(\boldsymbol{\beta}))=(\boldsymbol{\alpha},\boldsymbol{\beta})$．

证　充分性是显然的，取 $\boldsymbol{\alpha}=\boldsymbol{\beta}$，则有 $|f(\boldsymbol{\alpha})|^2=|\boldsymbol{\alpha}|^2$，所以 $|f(\boldsymbol{\alpha})|=|\boldsymbol{\alpha}|$．

反之，设 f 是一个正交变换，对于 $\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbf{V}$，有 $|f(\boldsymbol{\alpha}+\boldsymbol{\beta})|^2=|\boldsymbol{\alpha}+\boldsymbol{\beta}|^2$．

又$\because$ $|f(\boldsymbol{\alpha}+\boldsymbol{\beta})|^2=(f(\boldsymbol{\alpha}+\boldsymbol{\beta}),f(\boldsymbol{\alpha}+\boldsymbol{\beta}))=(f(\boldsymbol{\alpha})+f(\boldsymbol{\beta}),f(\boldsymbol{\alpha})+f(\boldsymbol{\beta}))=(f(\boldsymbol{\alpha}),f(\boldsymbol{\alpha}))+(f(\boldsymbol{\beta}),f(\boldsymbol{\beta}))+2(f(\boldsymbol{\alpha}),f(\boldsymbol{\beta}))$；而 $|\boldsymbol{\alpha}+\boldsymbol{\beta}|^2=(\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{\alpha}+\boldsymbol{\beta})=(\boldsymbol{\alpha},\boldsymbol{\alpha})+(\boldsymbol{\beta},\boldsymbol{\beta})+2(\boldsymbol{\alpha},\boldsymbol{\beta})$；由于 $(f(\boldsymbol{\alpha}),f(\boldsymbol{\alpha}))=(\boldsymbol{\alpha},\boldsymbol{\alpha})$；$(f(\boldsymbol{\beta}),f(\boldsymbol{\beta}))=(\boldsymbol{\beta},\boldsymbol{\beta})$．因此

$$(f(\boldsymbol{\alpha}),f(\boldsymbol{\beta}))=(\boldsymbol{\alpha},\boldsymbol{\beta})$$

由于两个非零向量的夹角是由它们的内积决定的，因此正交变换保持夹角不变．

定理 2.4.2　设 $\mathbf{V}$ 是 n 维欧氏空间，f 是 $\mathbf{V}$ 的一个线性变换．如果 f 是正交变换，则 f 将 $\mathbf{V}$ 的任意一个标准正交基变为 $\mathbf{V}$ 的一个标准正交基；反之，如果 f 将 $\mathbf{V}$ 的任意一个标准正交基变为 $\mathbf{V}$ 的一个标准正交基，则 f 是 $\mathbf{V}$ 的一个正交变换．

证　设 f 是 $\mathbf{V}$ 的一个正交变换，令 $\{\boldsymbol{\gamma}_1,\ \boldsymbol{\gamma}_2,\ \cdots,\ \boldsymbol{\gamma}_n\}$ 是 $\mathbf{V}$ 的任意一个标准正交基，由定理 2.4.1 得

$$(f(\boldsymbol{\gamma}_i),f(\boldsymbol{\gamma}_j))=(\boldsymbol{\gamma}_i,\boldsymbol{\gamma}_j)=\begin{cases}1, & i=j\\0, & i\neq j\end{cases}$$

$\therefore$ $\{f(\boldsymbol{\gamma}_1),f(\boldsymbol{\gamma}_2),\cdots,f(\boldsymbol{\gamma}_n)\}$ 也是 $\mathbf{V}$ 的一个标准正交基．

反之，设 $\mathbf{V}$ 的一个线性变换 f 将 $\mathbf{V}$ 的某一个标准正交基 $\{\boldsymbol{\gamma}_1,\ \boldsymbol{\gamma}_2,\ \cdots,\ \boldsymbol{\gamma}_n\}$ 变为标准正交基 $\{f(\boldsymbol{\gamma}_1),f(\boldsymbol{\gamma}_2),\cdots,f(\boldsymbol{\gamma}_n)\}$，令 $\boldsymbol{\alpha}=\sum\limits_{i=1}^{n}x_i\boldsymbol{\gamma}_i\in\mathbf{V}$，有

$$|f(\boldsymbol{\alpha})|^2=(f(\boldsymbol{\alpha}),f(\boldsymbol{\alpha}))=\left(\sum_{i=1}^{n}x_if(\boldsymbol{\gamma}_i),\sum_{j=1}^{n}x_jf(\boldsymbol{\gamma}_j)\right)$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_j(f(\boldsymbol{\gamma}_i),f(\boldsymbol{\gamma}_j))=\sum_{i=1}^{n}x_i^2=|\boldsymbol{\alpha}|^2$$

$\therefore f$ 是 $\boldsymbol{V}$ 的一个正交变换．

定义 2.4.2　如果实方阵 $\boldsymbol{U}$ 满足 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{E}$ 或 $\boldsymbol{U}^{-1}=\boldsymbol{U}^{\mathrm{T}}$，则称 $\boldsymbol{U}$ 为正交矩阵．

定理 2.4.3　实方阵 $\boldsymbol{U}$ 是正交矩阵的充分且必要条件为 $\boldsymbol{U}$ 的列向量是两两正交的单位向量.

定理 2.4.4　欧氏空间的线性变换是正交变换的充分且必要条件为它对于标准正交基的矩阵是正交矩阵．

证　设 $\boldsymbol{V}$ 是 n 维欧氏空间，f 是 $\boldsymbol{V}$ 的一个正交变换．

取定 $\boldsymbol{V}$ 的一个标准正交基$\{\boldsymbol{\gamma}_1，\boldsymbol{\gamma}_2，\cdots，\boldsymbol{\gamma}_n\}$，令 f 关于这个基的矩阵为 $\boldsymbol{U}=(u_{ij})$. 则有

$$f(\boldsymbol{\gamma}_j)=\sum_{i=1}^{n}u_{ij}\boldsymbol{\gamma}_i,\quad j=1,2,\cdots,n$$

$\because$ $\{\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_n\}$ 是 $\boldsymbol{V}$ 的一个标准正交基，

$$\therefore (f(\boldsymbol{\gamma}_i),f(\boldsymbol{\gamma}_j))=\left(\sum_{k=1}^{n}u_{ki}\boldsymbol{\gamma}_k,\sum_{l=1}^{n}u_{lj}\boldsymbol{\gamma}_l\right)=\sum_{k=1}^{n}\sum_{l=1}^{n}u_{ki}u_{lj}(\boldsymbol{\gamma}_k,\boldsymbol{\gamma}_l)=\sum_{k=1}^{n}u_{ki}u_{kj}\ .$$

由定理 2.4.2，有 $\sum\limits_{k=1}^{n}u_{ki}u_{kj}=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}$. 此式表明矩阵 $\boldsymbol{U}$ 满足 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{E}$.

因为 $\boldsymbol{U}$ 是可逆矩阵，所以 $\boldsymbol{U}^{-1}=\boldsymbol{U}^{\mathrm{T}}$，即 $\boldsymbol{U}$ 是一个正交矩阵．

反之，若 n 维欧氏空间 $\boldsymbol{V}$ 的一个线性变换 f 关于某一个标准正交基 $\{\boldsymbol{\gamma}_1，\boldsymbol{\gamma}_2，\cdots，\boldsymbol{\gamma}_n\}$ 的矩阵 $\boldsymbol{U}=(u_{ij})$ 是一个正交矩阵，则

$$f(\boldsymbol{\gamma}_j)=\sum_{i=1}^{n}u_{ij}\boldsymbol{\gamma}_i,\quad j=1,2,\cdots,n$$

并且有 $(f(\boldsymbol{\gamma}_i)，f(\boldsymbol{\gamma}_j))=\sum\limits_{k=1}^{n}u_{ki}u_{kj}=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}$.

所以 $\{f(\boldsymbol{\gamma}_1),f(\boldsymbol{\gamma}_2),\cdots,f(\boldsymbol{\gamma}_n)\}$ 是 $\boldsymbol{V}$ 的一个标准正交基，由定理 2.4.2，f 是 $\boldsymbol{V}$ 的一个正交变换．

推论 2.4.1　正交矩阵是可逆矩阵．

证　设 $\boldsymbol{U}$ 是任意一个正交矩阵，则有 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{E}$，两侧取行列式，即

$$\det(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})=\det\boldsymbol{U}^{\mathrm{T}}\det\boldsymbol{U}=(\det\boldsymbol{U})^2=\det\boldsymbol{E}=1$$

所以 $\det\boldsymbol{U}=\pm1\neq0$，因此 $\boldsymbol{U}$ 是可逆矩阵．

推论 2.4.2　正交矩阵的逆矩阵也是正交矩阵．

证　设 $\boldsymbol{U}$ 是任意一个正交矩阵，

$\because (\boldsymbol{U}^{-1})^{\mathrm{T}}\boldsymbol{U}^{-1}=(\boldsymbol{U}^{\mathrm{T}})^{-1}\boldsymbol{U}^{-1}=(\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}})^{-1}=\boldsymbol{E}$,

$\therefore \boldsymbol{U}^{-1}$ 也是一个正交矩阵．

推论 2.4.3　两个正交矩阵的乘积仍是正交矩阵．

证　设 $\boldsymbol{U}_1$，$\boldsymbol{U}_2$ 是任意两个正交矩阵，则

$$(\boldsymbol{U}_1\boldsymbol{U}_2)^{\mathrm{T}}\boldsymbol{U}_1\boldsymbol{U}_2=\boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{U}_1\boldsymbol{U}_2=\boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{U}_2=\boldsymbol{E}$$

所以结论成立．

§2.5 对称变换与对称矩阵

欧氏空间的另一个重要的线性变换是对称变换，本节中我们将讨论有限维欧氏空间中对称变换的一些性质．

定义 2.5.1 设 f 是 n 维欧氏空间 $\boldsymbol{V}$ 的一个线性变换，如果对于任意的向量 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\boldsymbol{V}$，有 $(f(\boldsymbol{\alpha}),\boldsymbol{\beta})=(\boldsymbol{\alpha},f(\boldsymbol{\beta}))$ 成立，则称 f 是一个对称变换．

定理 2.5.1 如果 f 是 n 维欧氏空间 $\boldsymbol{V}$ 的一个对称变换，$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$ 是 $\boldsymbol{V}$ 的任意一个标准正交基，$\boldsymbol{A}=(a_{ij})$ 是 f 关于这个基的矩阵，那么 $\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$.

证 由于 $f(\boldsymbol{\alpha}_j)=\sum\limits_{k=1}^{n}a_{kj}\boldsymbol{\alpha}_k$，$1\leqslant j\leqslant n$. 又因为 f 是对称变换，$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$ 是标准正交基，所以

$$a_{ji}=\left(\sum_{k=1}^{n}a_{ki}\boldsymbol{\alpha}_k,\boldsymbol{\alpha}_j\right)=(f(\boldsymbol{\alpha}_i),\boldsymbol{\alpha}_j)=(\boldsymbol{\alpha}_i,f(\boldsymbol{\alpha}_j))=\left(\boldsymbol{\alpha}_i,\sum_{k=1}^{n}a_{kj}\boldsymbol{\alpha}_k\right)=a_{ij}$$

$$\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$$

而我们知道满足 $\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$ 的矩阵是对称矩阵，所以定理 2.5.1 等于是说 n 维欧氏空间 $\boldsymbol{V}$ 的对称变换关于任意一个标准正交基的矩阵是一个对称矩阵，反之有如下的定理．

定理 2.5.2 设 f 是 n 维欧氏空间 $\boldsymbol{V}$ 的一个线性变换，如果 f 关于 $\boldsymbol{V}$ 的一个标准正交基的矩阵是对称矩阵，则 f 是一个对称变换．

证 设 f 关于 $\boldsymbol{V}$ 的一个标准正交基 $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$ 的矩阵 $\boldsymbol{A}=(a_{ij})$ 是对称矩阵，对任意的 $\boldsymbol{\alpha}=\sum\limits_{i=1}^{n}x_i\boldsymbol{\alpha}_i$，$\boldsymbol{\beta}=\sum\limits_{i=1}^{n}y_i\boldsymbol{\alpha}_i\in V$，则

$$\begin{aligned}(f(\boldsymbol{\alpha}),\boldsymbol{\beta})&=\left(\sum_{i=1}^{n}x_if(\boldsymbol{\alpha}_i),\sum_{j=1}^{n}y_j\boldsymbol{\alpha}_j\right)=\left(\sum_{i=1}^{n}x_i\left(\sum_{k=1}^{n}a_{ki}\boldsymbol{\alpha}_k\right),\sum_{j=1}^{n}y_j\boldsymbol{\alpha}_j\right)\\&=\left(\sum_{k=1}^{n}\left(\sum_{i=1}^{n}a_{ki}x_i\right)\boldsymbol{\alpha}_k,\sum_{j=1}^{n}y_j\boldsymbol{\alpha}_j\right)=\sum_{j=1}^{n}\sum_{i=1}^{n}a_{ji}x_iy_j\end{aligned}$$

同样有

$$(\boldsymbol{\alpha},f(\boldsymbol{\beta}))=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_iy_j$$

因为 $a_{ij}=a_{ji}$，所以 $(f(\boldsymbol{\alpha}),\boldsymbol{\beta})=(\boldsymbol{\alpha},f(\boldsymbol{\beta}))$，即 f 是一个对称变换．

对称变换有下列几个基本性质．

定理 2.5.3 设 f 是 n 维欧氏空间 $\boldsymbol{V}$ 的一个对称变换，则有：

（1）实对称矩阵的特征根是实数；

（2）f 的属于不同特征根的特征向量彼此正交．

证 （1）设 $\boldsymbol{A}=(a_{ij})$ 是一个 n 阶实对称矩阵，λ 是 $\boldsymbol{A}$ 在复数域 $\mathbf{C}$ 中的一个特征根，因此存在不全为零的复数 c_1，c_2，$\cdots$，c_n，使得

$$\boldsymbol{A}\begin{pmatrix}c_1\\c_2\\\vdots\\c_n\end{pmatrix}=\lambda\begin{pmatrix}c_1\\c_2\\\vdots\\c_n\end{pmatrix}$$

以 $\overline{c_i}$ 表示 c_i 的共轭复数，则

$$(\overline{c_1},\overline{c_2},\cdots,\overline{c_n})\boldsymbol{A}\begin{pmatrix}c_1\\c_2\\\vdots\\c_n\end{pmatrix}=\lambda(\overline{c_1},\overline{c_2},\cdots,\overline{c_n})\begin{pmatrix}c_1\\c_2\\\vdots\\c_n\end{pmatrix}$$

即

$$\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\,\overline{c_i}c_j=\lambda\sum_{i=1}^{n}\overline{c_i}c_i$$

对此式两侧取共轭复数有

$$\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}c_i\,\overline{c_j}=\overline{\lambda}\sum_{i=1}^{n}c_i\,\overline{c_i}$$

又因为 $a_{ij}=a_{ji}$，所以 $\lambda\sum_{i=1}^{n}c_i\,\overline{c_i}=\overline{\lambda}\sum_{i=1}^{n}c_i\,\overline{c_i}$，而 c_i 不全为零，由此 $\sum_{i=1}^{n}c_i\,\overline{c_i}$ 是一个正实数，也就是说 $\lambda=\overline{\lambda}\in\mathbf{R}$.

(2) 如果 λ，μ 是 f 的特征根，并且 $\lambda\neq\mu$，令 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 分别为对应于 λ 和 μ 的特征向量，

$\therefore f(\boldsymbol{\alpha})=\lambda\boldsymbol{\alpha},f(\boldsymbol{\beta})=\mu\boldsymbol{\beta}$,

$\therefore \lambda(\boldsymbol{\alpha},\boldsymbol{\beta})=(\lambda\boldsymbol{\alpha},\boldsymbol{\beta})=(f(\boldsymbol{\alpha}),\boldsymbol{\beta})=(\boldsymbol{\alpha},f(\boldsymbol{\beta}))=(\boldsymbol{\alpha},\mu\boldsymbol{\beta})=\mu(\boldsymbol{\alpha},\boldsymbol{\beta})$,

$\because \lambda\neq\mu$,所以$(\boldsymbol{\alpha},\boldsymbol{\beta})=0$.

同时我们还得到下面定理.

定理 2.5.4 设 f 是 n 维欧氏空间 $\mathbf{V}$ 的一个对称变换，则有：

(1) 存在 $\mathbf{V}$ 的一个标准正交基，f 关于这个基的矩阵是对角矩阵；

(2) 如果 $\boldsymbol{A}=(a_{ij})$ 是一个 n 阶实对称矩阵，则存在着一个 n 阶正交矩阵 $\boldsymbol{U}$，使得 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{U}$ 是对角矩阵.

【例 2.5.1】 给定欧氏空间 $\mathbf{V}$ 的变换 $\boldsymbol{T}(\boldsymbol{\alpha})=\boldsymbol{\alpha}-2(\boldsymbol{\alpha},\boldsymbol{\alpha}_0)\boldsymbol{\alpha}_0$，$\forall\boldsymbol{\alpha}\in\mathbf{V}$，$\boldsymbol{\alpha}_0\in\mathbf{V}$ 是取定的单位元素

证明：(1) $\boldsymbol{T}$ 是线性变换；

(2) $\boldsymbol{T}$ 既是正交变换，又是对称变换.

证 (1) $\forall\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\mathbf{V}$，$k$，$l\in\mathbf{R}$，有

$$\begin{aligned}T(k\boldsymbol{\alpha},l\boldsymbol{\beta})&=(k\boldsymbol{\alpha}+l\boldsymbol{\beta})-2(k\boldsymbol{\alpha}+l\boldsymbol{\beta},\boldsymbol{\alpha}_0)\boldsymbol{\alpha}_0\\&=k\boldsymbol{\alpha}+l\boldsymbol{\beta}-2k(\boldsymbol{\alpha},\boldsymbol{\alpha}_0)\boldsymbol{\alpha}_0-2l(\boldsymbol{\beta},\boldsymbol{\alpha}_0)\boldsymbol{\alpha}_0-k[\boldsymbol{\alpha}-2(\boldsymbol{\alpha},\boldsymbol{\alpha}_0)\boldsymbol{\alpha}_0]+l[\boldsymbol{\beta}-2(\boldsymbol{\beta},\boldsymbol{\alpha}_0)\boldsymbol{\alpha}_0]\\&=k\boldsymbol{T}(\boldsymbol{\alpha})+l\boldsymbol{T}(\boldsymbol{\beta})\end{aligned}$$

故 T 是线性变换.

(2) $\forall\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\mathbf{V}$，有

$$\begin{aligned}(\boldsymbol{T}(\boldsymbol{\alpha}),\boldsymbol{T}(\boldsymbol{\beta}))&=(\boldsymbol{\alpha}-2(\boldsymbol{\alpha},\boldsymbol{\alpha}_0)\boldsymbol{\alpha}_0,\boldsymbol{\beta}-2(\boldsymbol{\beta},\boldsymbol{\alpha}_0)\boldsymbol{\alpha}_0)\\&=(\boldsymbol{\alpha},\boldsymbol{\beta})-2(\boldsymbol{\beta},\boldsymbol{\alpha}_0)(\boldsymbol{\alpha},\boldsymbol{\alpha}_0)-2(\boldsymbol{\alpha},\boldsymbol{\alpha}_0)(\boldsymbol{\alpha}_0,\boldsymbol{\beta})+4(\boldsymbol{\alpha},\boldsymbol{\alpha}_0)(\boldsymbol{\beta},\boldsymbol{\alpha}_0)(\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_0)\\&=(\boldsymbol{\alpha},\boldsymbol{\beta})-4(\boldsymbol{\beta},\boldsymbol{\alpha}_0)(\boldsymbol{\alpha},\boldsymbol{\alpha}_0)+4(\boldsymbol{\alpha},\boldsymbol{\alpha}_0)(\boldsymbol{\beta},\boldsymbol{\alpha}_0)=(\boldsymbol{\alpha},\boldsymbol{\beta})\end{aligned}$$

从而 $\boldsymbol{T}$ 是正交变换；

又 $(\boldsymbol{T}(\boldsymbol{\alpha}),\boldsymbol{\beta})=(\boldsymbol{\alpha}-2(\boldsymbol{\alpha},\boldsymbol{\alpha}_0)\boldsymbol{\alpha}_0,\boldsymbol{\beta})=(\boldsymbol{\alpha},\boldsymbol{\beta})-2(\boldsymbol{\alpha},\boldsymbol{\alpha}_0)(\boldsymbol{\alpha}_0,\boldsymbol{\beta})$，

而 $(\boldsymbol{\alpha},\boldsymbol{T}(\boldsymbol{\beta}))=(\boldsymbol{\alpha},\boldsymbol{\beta}-2(\boldsymbol{\beta},\boldsymbol{\alpha}_0)\boldsymbol{\alpha}_0)=(\boldsymbol{\alpha},\boldsymbol{\beta})-2(\boldsymbol{\beta},\boldsymbol{\alpha}_0)(\boldsymbol{\alpha},\boldsymbol{\alpha}_0)$，

则 $(T(\boldsymbol{\alpha}),\boldsymbol{\beta})=(\boldsymbol{\alpha},T(\boldsymbol{\beta}))$，$\therefore T$ 是对称变换．

【例 2.5.2】 已知 $\mathbf{R}^{2\times2}$ 的线性变换

$$T\begin{pmatrix}x_1 & x_2\\x_3 & x_4\end{pmatrix}=\begin{pmatrix}x_2+x_3-x_4 & x_1-x_3+x_4\\x_1-x_2+x_4 & -x_1+x_2+x_3\end{pmatrix},\ \forall\begin{pmatrix}x_1 & x_2\\x_3 & x_4\end{pmatrix}\in\mathbf{R}^{2\times2}$$

(1) 证明 T 是对称变换；

(2) 求 $\mathbf{R}^{2\times2}$ 的一组标准正交基，使 T 在该基下的矩阵为对角矩阵．

解 (1) 取 $\mathbf{R}^{2\times2}$ 的标准正交基 E_{11}，E_{12}，E_{21}，E_{22}，可求得 T 在该基下的矩阵为

$$A=\begin{pmatrix}0 & 1 & 1 & -1\\1 & 0 & -1 & 1\\1 & -1 & 0 & 1\\-1 & 1 & 1 & 0\end{pmatrix}$$

由于 A 是实对称矩阵，所以 T 是对称变换；

(2) 因为 $\det(\lambda E-A)=(\lambda-1)^3(\lambda+3)$，所以 A 的特征值为 $\lambda_1=\lambda_2=\lambda_3=1$，$\lambda_4=-3$，可求得对应于 $\lambda_1=\lambda_2=\lambda_3=1$ 的特征向量为

$$\boldsymbol{p}_1=(1,1,0,0)^{\mathrm{T}},\quad \boldsymbol{p}_2=(1,0,1,0)^{\mathrm{T}},\quad \boldsymbol{p}_2=(-1,0,0,1)^{\mathrm{T}}$$

将其正交化并单位化得

$$\boldsymbol{q}_1=\left(\frac{1}{\sqrt2},\frac{1}{\sqrt2},0,0\right)^{\mathrm{T}},\quad \boldsymbol{q}_2=\left(\frac{1}{\sqrt6},\frac{-1}{\sqrt6},\frac{2}{\sqrt6},0\right)^{\mathrm{T}},\quad \boldsymbol{q}_3=\left(\frac{-1}{2\sqrt3},\frac{1}{2\sqrt3},\frac{1}{2\sqrt3},\frac{3}{2\sqrt3}\right)^{\mathrm{T}}$$

又可求得对应于 $\lambda_4=-3$ 的特征向量为

$$\boldsymbol{p}_4=(1,-1,-1,1)^{\mathrm{T}}$$

将其正交化并单位化得

$$\boldsymbol{q}_2=\left(\frac{1}{2},\frac{-1}{2}\frac{-1}{2},\frac{1}{2}\right)^{\mathrm{T}}$$

故正交矩阵

$$Q=\begin{pmatrix}\frac{1}{\sqrt2} & \frac{1}{\sqrt6} & \frac{-1}{2\sqrt3} & \frac{1}{2}\\ \frac{1}{\sqrt2} & \frac{-1}{\sqrt6} & \frac{1}{2\sqrt3} & \frac{-1}{2}\\ 0 & \frac{2}{\sqrt6} & \frac{1}{2\sqrt3} & -\frac{1}{2}\\ 0 & 0 & \frac{3}{2\sqrt3} & \frac{1}{2}\end{pmatrix}$$

使得

$$Q^{-1}AQ=\Lambda=\begin{pmatrix}1 & & & \\ & 1 & & \\ & & 1 & \\ & & & -3\end{pmatrix}$$

由 $(X_1,X_2,X_3,X_4)=(E_{11},E_{12},E_{21},E_{22})Q$ 求得 $\mathbf{R}^{2\times2}$ 的标准正交基

$$X_1=\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ 0 & 0\end{pmatrix},X_2=\begin{pmatrix}\frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{6}}\\ \frac{2}{\sqrt{6}} & 0\end{pmatrix},X_3=\begin{pmatrix}\frac{-1}{2\sqrt{3}} & \frac{1}{2\sqrt{3}}\\ \frac{1}{2\sqrt{3}} & \frac{3}{2\sqrt{3}}\end{pmatrix},X_4=\begin{pmatrix}\frac{1}{2} & \frac{-1}{2}\\ \frac{-1}{2} & \frac{1}{2}\end{pmatrix}$$

$\boldsymbol{T}$ 在该基下的矩阵为 $\boldsymbol{A}$.

§2.6 特征值与特征向量

从前面的讨论可知，在有限维线性空间 $\boldsymbol{V}$ 中，取定一个基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_s$ 后，线性变换 f 与矩阵 $\boldsymbol{A}$ 之间存在着一一对应关系. 因此，利用矩阵来研究线性变换十分方便.

每个可逆矩阵都是基变换矩阵，而每个基变换矩阵也都是可逆矩阵. 如果给定 $\boldsymbol{V}$ 的一个基与一个线性变换 f，且与 f 对应的矩阵为 $\boldsymbol{A}$，则与 $\boldsymbol{A}$ 相似的所有矩阵恰好是 f 在 $\boldsymbol{V}$ 的不同的基下的矩阵.

那么对于每一个给定的线性变换，如何适当选择 $\boldsymbol{V}$ 的一个基，使得该线性变换在此基下的矩阵最为简单呢? 这是本节结尾要讨论的问题. 为此，我们引入特征值与特征向量的概念.

特征值与特征向量的概念在实践中也有着广泛的应用，大型建筑物与机械的振动，机翼的颤振以及调节系统的自振等都是常见的例子.

定义 2.6.1 设 f 是数域 P 上 n 维线性空间 $\boldsymbol{V}$ 的线性变换，即 $f\in \boldsymbol{L}(\boldsymbol{V})$. 如果对于数域 P 中的某一个数 λ_0，存在非零向量 $\boldsymbol{\alpha}\in\boldsymbol{V}$，使得 $f(\boldsymbol{\alpha})=\lambda_0\boldsymbol{\alpha}$ 成立，则称 λ_0 为线性变换 f 的特征值，$\boldsymbol{\alpha}$ 称为线性变换 f 的属于特征值 λ_0 的特征向量.

显然，如果 $\boldsymbol{\alpha}$ 是 f 的属于特征值 λ_0 的一个特征向量，则对于任意的数 $k\in P$，都有

$$f(k\boldsymbol{\alpha})=kf(\boldsymbol{\alpha})=k\lambda_0\boldsymbol{\alpha}=\lambda_0(k\boldsymbol{\alpha}) \tag{2.6.1}$$

因此，如果 $\boldsymbol{\alpha}$ 是 f 的一个特征向量，则由 $\boldsymbol{\alpha}$ 所生成的一维子空间 $\boldsymbol{U}=\{k\boldsymbol{\alpha}\,|\,k\in P\}=\mathrm{Span}\{\boldsymbol{\alpha}\}$ 在 f 之下不变；反之，如果 $\boldsymbol{V}$ 的一个一维子空间 $\boldsymbol{U}$ 在 f 之下不变，则 $\boldsymbol{U}$ 中每一个非零向量都是 f 的属于同一特征值的特征向量. 所以一维不变子空间与特征值之间有着密切的关系.

而且，由式 (2.6.1) 可见，如果 $\boldsymbol{\alpha}$ 是 f 的属于特征值 λ_0 的特征向量，$k\neq 0$，那么 $k\boldsymbol{\alpha}$ 也是属于特征值 λ_0 的特征向量，即 $f(k\boldsymbol{\alpha})=\lambda_0(k\boldsymbol{\alpha})$，$k\in P$. 由此可见，特征向量 $\boldsymbol{\alpha}$ 并非由特征值 λ_0 唯一确定；但是特征值却被特征向量唯一确定.

【例 2.6.1】 令 $\boldsymbol{D}$ 表示定义在实数域 $\boldsymbol{R}$ 上的可微分任意次的实函数所构成的线性空间，$\delta: f(x)\mapsto f'(x)$ 为求导数的运算，δ 是 $\boldsymbol{D}$ 的一个线性变换.

对每一个 $\lambda\in\boldsymbol{R}$，有 $\delta(\mathrm{e}^{\lambda x})=\lambda\mathrm{e}^{\lambda x}$，所以对任意的 $\lambda\in\boldsymbol{R}$，都是 δ 的特征根，而 $\mathrm{e}^{\lambda x}$ 是属于 λ 的一个特征向量.

【例 2.6.2】 设 $\boldsymbol{P}[x]$ 是数域 P 上一切一元多项式所构成的线性空间，$\sigma: f(x)\mapsto xf(x)$，易证明 σ 是 $\boldsymbol{P}[x]$ 的一个线性变换.

比较变换前后的多项式的次数可知，对任意的 $\lambda\in P$，均不存在非零多项式 $f(x)$，使得 $xf(x)=\lambda f(x)$，所以 σ 无特征值.

现在我们来讨论线性变换 f 的特征值与其所对应的矩阵的特征值之间的关系.

设 $\boldsymbol{V}$ 是数域 P 上 n 维线性空间，取定 $\boldsymbol{V}$ 一个基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$，令 $\boldsymbol{V}$ 的线性变换 f 关

于这个基的矩阵为$A=(a_{ij})$.

如果$\alpha=x_1\alpha_1+x_2\alpha_2+\cdots+x_n\alpha_n$是$f$的属于特征值$\lambda$的一个特征向量，由定义2.6.1有

$$A\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\lambda\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\quad\text{或}\quad(\lambda E-A)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix}\tag{2.6.2}$$

因为$\alpha\neq0$所以方程组（2.6.2）有非零解，所以

$$|\lambda E-A|=\begin{vmatrix}\lambda-a_{11}&-a_{12}&\cdots&-a_{1n}\\-a_{21}&\lambda-a_{22}&\cdots&-a_{2n}\\\cdots&\cdots&\cdots&\cdots\\-a_{n1}&-a_{n2}&\cdots&\lambda-a_{nn}\end{vmatrix}=0\tag{2.6.3}$$

反之，如果$\lambda\in P$满足式（2.6.3），则式（2.6.2）有非零解x_1，x_2，…，x_n. 则$\alpha=x_1\alpha_1+x_2\alpha_2+\cdots+x_n\alpha_n$满足定义式$f(\alpha)=\lambda\alpha$，因此$\lambda$是$f$的一个特征值.

式（2.6.3）中的行列式十分重要，引入以下的定义.

定义2.6.2 设矩阵$A=(a_{ij})_{n\times n}$，$a_{ij}\in P$，矩阵$\lambda E-A$为A的特征矩阵，$\det(\lambda E-A)$为A的特征多项式，$\det(\lambda E-A)=0$为A的特征方程，其根称为A的特征值（或特征根）. $\det(\lambda E-A)$的展开式为λ的n次多项式，记作$f_A(\lambda)=\lambda^n+a_1\lambda^{n-1}+\cdots+a_{n-1}\lambda+a_n$.

对于A的特征多项式

$$f_A(x)=\begin{vmatrix}x-a_{11}&-a_{12}&\cdots&-a_{1n}\\-a_{21}&x-a_{22}&\cdots&-a_{2n}\\\cdots&\cdots&\cdots&\cdots\\-a_{n1}&-a_{n2}&\cdots&x-a_{nn}\end{vmatrix}$$

将这个行列式展开，得到$P_n[x]$中的一个多项式，这个多项式的最高次项为x^n，它出现在行列式主对角线上的元素的乘积中，即

$$(x-a_{11})(x-a_{22})\cdots(x-a_{nn})\tag{2.6.4}$$

而行列式的其余各项至多含有$n-2$个主对角线上的元素.

因此，$f_A(x)$是式（2.6.4）和一个至多是x的$n-2$次多项式的和，所以$f_A(x)$中次数大于$n-2$的项只出现在式（2.6.4）中. 即

$$f_A(x)=x^n-(a_{11}+a_{22}+\cdots+a_{nn})x^{n-1}+\cdots$$

其中$-(a_{11}+a_{22}+\cdots+a_{nn})$乘以$(-1)$就是矩阵$A$的主对角线上的元素之和，我们称之为矩阵$A$的迹，记作$\operatorname{tr}(A)$，即$\operatorname{tr}(A)=(a_{11}+a_{22}+\cdots+a_{nn})$.

定义2.6.3 在n阶方阵$A=(a_{ij})$中，任意选取行与列次相同的k行与k列$(1\leqslant k\leqslant n)$，位于这$k$行与$k$列交叉处的$k^2$个元素按原有位置组成的$k$阶行列式之和，叫做矩阵$A$的$k$阶迹，记作$\operatorname{tr}^{[k]}(A)$，即

$$\operatorname{tr}^{[k]}(A)=\sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}\begin{vmatrix}a_{i_1i_1}&\cdots&a_{i_1i_k}\\\cdots&\cdots&\cdots\\a_{i_ki_1}&\cdots&a_{i_ki_k}\end{vmatrix},\quad(k=1,2,\cdots,n)$$

式中：$a_{i_ti_s}$ $(1\leqslant i_t,\ i_s\leqslant n;\ t,\ s=1,\ 2,\ \cdots,\ n;\ i_1<i_2<\cdots<i_k)$是位于矩阵$\boldsymbol{A}$中第$i_t$行、$i_s$列交叉处的元素；$n$阶方阵$\boldsymbol{A}$的$k$阶迹$\mathrm{tr}^{[k]}(\boldsymbol{A})$为$C_n^k$个$k$阶行列式之和 $(k=1,\ 2,\ \cdots,\ n)$.

根据这个定义，显然有

$$\mathrm{tr}^{[1]}(\boldsymbol{A})=\mathrm{tr}(\boldsymbol{A})=(a_{11}+a_{22}+\cdots+a_{nn})$$

$$\mathrm{tr}^{[2]}(\boldsymbol{A})=\sum_{1\leqslant i<j\leqslant n}\begin{vmatrix}a_{ii} & a_{ij}\\ a_{ji} & a_{jj}\end{vmatrix}=\begin{vmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{vmatrix}+\begin{vmatrix}a_{11} & a_{13}\\ a_{31} & a_{33}\end{vmatrix}+\cdots+\begin{vmatrix}a_{11} & a_{1n}\\ a_{n1} & a_{nn}\end{vmatrix}+\cdots+\begin{vmatrix}a_{n-1n-1} & a_{n-1n}\\ a_{nn-1} & a_{nn}\end{vmatrix}$$

$$\cdots\cdots$$

$$\mathrm{tr}^{[n]}(\boldsymbol{A})=|\boldsymbol{A}|$$

又因为，$f_A(0)=(-1)^n|\boldsymbol{A}|$，所以$f_A(x)$的常数项为$(-1)^n|\boldsymbol{A}|$.

定理 2.6.1 n阶方阵$\boldsymbol{A}$的特征多项式$f_A(\lambda)=|\lambda\boldsymbol{I}-\boldsymbol{A}|$可以写成如下形式

$$f_A(\lambda)=\sum_{k=0}^{n}(-1)^k\mathrm{tr}^{[k]}(\boldsymbol{A})\lambda^{n-k} \tag{2.6.5}$$

其中

$$\mathrm{tr}^{[0]}(\boldsymbol{A})=1,\mathrm{tr}^{[k]}(\boldsymbol{A})=\sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}\begin{vmatrix}a_{i_1i_1} & \cdots & a_{i_1i_k}\\ \cdots & \cdots & \cdots\\ a_{i_ki_1} & \cdots & a_{i_ki_k}\end{vmatrix}\quad(k=1,2,\cdots,n)$$

【例 2.6.3】 设矩阵$\boldsymbol{A}=\begin{pmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{pmatrix}$，试计算$\boldsymbol{A}$的特征多项式$f_A(x)$.

解 $f_A(x)=\begin{vmatrix}x-a_{11} & -a_{12} & -a_{13}\\ -a_{21} & x-a_{22} & -a_{23}\\ -a_{31} & -a_{32} & x-a_{33}\end{vmatrix}=x^3-\mathrm{tr}^{[1]}(\boldsymbol{A})x^2+\mathrm{tr}^{[2]}(\boldsymbol{A})x-|\boldsymbol{A}|$.

【例 2.6.4】 设线性变换$f\in L(\mathbf{R}^3)$，f在基$\{\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \boldsymbol{\alpha}_3\}$下的矩阵为$\boldsymbol{A}=\begin{pmatrix}5 & 6 & -3\\ -1 & 0 & 1\\ 1 & 2 & -1\end{pmatrix}$，试求$f$的特征值与特征向量.

解 $\because$ $f_A(\lambda)=|\lambda\boldsymbol{E}-\boldsymbol{A}|=\lambda^3-\mathrm{tr}^{[1]}(\boldsymbol{A})\lambda^2+\mathrm{tr}^{[2]}(\boldsymbol{A})\lambda-\mathrm{tr}^{[3]}(\boldsymbol{A})$

$$=\lambda^3-(5+0-1)\lambda^2+\left(\begin{vmatrix}5 & 6\\ -1 & 0\end{vmatrix}+\begin{vmatrix}5 & -3\\ 1 & -1\end{vmatrix}+\begin{vmatrix}0 & 1\\ 2 & -1\end{vmatrix}\right)\lambda-|\boldsymbol{A}|$$

$$=(\lambda-2)(\lambda^2-2\lambda-2)$$

令$f_A(\lambda)=0$，得到f的三个特征值$\lambda_1=2$，$\lambda_2=1+\sqrt{3}$，$\lambda_3=1-\sqrt{3}$.

当$\lambda_1=2$时，解方程组$(2\boldsymbol{E}-\boldsymbol{A})x=0$，得$\begin{cases}x_1=-2x_2\\ x_3=0\end{cases}$，其基础解系为$(-2,\ 1,\ 0)^{\mathrm{T}}$，所以$f$的对应于特征值$\lambda_1=2$的一切特征向量为$k(-2\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2)$，$k\in\mathbf{C}$；

当$\lambda_2=1+\sqrt{3}$时，解方程组$[(1+\sqrt{3})\boldsymbol{E}-\boldsymbol{A}]x=0$，得$\begin{cases}x_1=-3x_2\\ x_3=(-2+\sqrt{3})x_2\end{cases}$，其基础解系为$(3,\ -1,\ 2-\sqrt{3})^{\mathrm{T}}$，所以$f$的对应于特征值$\lambda_2=1+\sqrt{3}$的一切特征向量为$k\ [3\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2+(2-\sqrt{3})\boldsymbol{\alpha}_3]$，$k\in\mathbf{C}$；

当$\lambda_3=1-\sqrt{3}$时，解方程组$[(1-\sqrt{3})\boldsymbol{E}-\boldsymbol{A}]x=0$，得$\begin{cases}x_1=-3x_2\\x_3=-(2+\sqrt{3})x_2\end{cases}$，其基础解系为$(3,\ -1,\ 2+\sqrt{3})^{\mathrm{T}}$，所以 f 的对应于特征值 $\lambda_3=1-\sqrt{3}$ 的一切特征向量为 $k[3\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2+(2+\sqrt{3})\boldsymbol{\alpha}_3]$，$k\in\mathbf{C}$.

【例 2.6.5】 已知 $\mathbf{R}^{2\times2}$的线性变换

$$\boldsymbol{T}(\boldsymbol{X})=\boldsymbol{MXN},\left[\forall\boldsymbol{X}\in\mathbf{R}^{2\times2},\boldsymbol{M}=\begin{pmatrix}1&0\\1&1\end{pmatrix},\boldsymbol{N}=\begin{pmatrix}1&-1\\-1&1\end{pmatrix}\right]$$

求 $\boldsymbol{T}$ 的特征值与特征向量.

解 取 $\mathbf{R}^{2\times2}$的标准正交基 $\boldsymbol{E}_{11}$，$\boldsymbol{E}_{12}$，$\boldsymbol{E}_{21}$，$\boldsymbol{E}_{22}$，可求得 $\boldsymbol{T}$ 在该基下的矩阵为

$$\boldsymbol{A}=\begin{pmatrix}1&-1&0\\-1&1&0\\1&-1&1\\-1&1&-1\end{pmatrix}$$

因为 $\det(\lambda\boldsymbol{E}-\boldsymbol{A})=\lambda^2(\lambda-2)^2$，所以 $\boldsymbol{A}$ 的特征值为$\lambda_1=\lambda_2=0$，$\lambda_3=\lambda_4=2$，可求得对应于$\lambda_1=\lambda_2=0$ 的特征向量为

$$\boldsymbol{p}_1=(1,1,0,0)^{\mathrm{T}},\boldsymbol{p}_2=(0,0,1,1)^{\mathrm{T}}$$

对应于 $\lambda_3=\lambda_4=2$ 的特征向量为

$$\boldsymbol{p}_3=(0,0,-1,1)^{\mathrm{T}}$$

故线性变换 T 的特征值为

$$\lambda_1=\lambda_2=0,\lambda_3=\lambda_4=2$$

对应特征值 $\lambda_1=\lambda_2=0$ 的线性无关特征向量为

$$\boldsymbol{X}_1=\boldsymbol{E}_{11}+\boldsymbol{E}_{12}=\begin{pmatrix}1&1\\0&0\end{pmatrix},\boldsymbol{X}_2=\boldsymbol{E}_{21}+\boldsymbol{E}_{22}=\begin{pmatrix}0&0\\1&1\end{pmatrix}$$

全部特征向量为 $k_1\boldsymbol{X}_1+k_2\boldsymbol{X}_2,k_1,k_2$ 不全为零.

对应特征值 $\lambda_3=\lambda_4=2$ 的线性无关特征向量为

$$\boldsymbol{X}_3=-\boldsymbol{E}_{21}+\boldsymbol{E}_{22}=\begin{pmatrix}0&0\\-1&1\end{pmatrix}$$

全部特征向量为

$$k\boldsymbol{X}_3\,(k\neq0)$$

现在我们来讨论矩阵的特征值与特征向量之间的关系.

定义 2.6.4 设线性变换 $f\in L(\boldsymbol{V})$，对 f 的任意一个特征值λ_0，满足 $f(\boldsymbol{\alpha})=\lambda_0\boldsymbol{\alpha}$ 的向量$\boldsymbol{\alpha}$ 组成的集合叫做f 的一个特征子空间，记作 $\boldsymbol{V}_{\lambda_0}=\{\boldsymbol{\alpha}\,|\,f(\boldsymbol{\alpha})=\lambda_0\boldsymbol{\alpha},\ \boldsymbol{\alpha}\in\boldsymbol{V}\}$.

显然特征子空间 $\boldsymbol{V}_{\lambda_0}$的维数 $\dim\boldsymbol{V}_{\lambda_0}$等于属于特征值$\lambda_0$ 的线性无关的特征向量的个数，我们称其为特征值λ_0 的**几何重数**；而特征值 λ_0 在特征多项式 $f(\lambda)$中的根的重数称为特征值λ_0 的**代数重数**.

定理 2.6.2 设线性变换 $f\in L(\boldsymbol{V})$，则 f 的任意特征值的几何重数不大于它的代数重数.

证 取特征子空间 $\boldsymbol{V}_{\lambda_0}$的一个基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_k$，则 $\dim\boldsymbol{V}_{\lambda_0}=k>0$.

将其扩充为 V 的一个基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_k$，$\boldsymbol{\alpha}_{k+1}$，…，$\boldsymbol{\alpha}_n$.
则有

$$\begin{aligned}
f(\boldsymbol{\alpha}_1) &= \lambda_0\boldsymbol{\alpha}_1 \\
f(\boldsymbol{\alpha}_2) &= \lambda_0\boldsymbol{\alpha}_2 \\
&\vdots \\
f(\boldsymbol{\alpha}_k) &= \lambda_0\boldsymbol{\alpha}_k \\
f(\boldsymbol{\alpha}_{k+1}) &= a_{1,k+1}\boldsymbol{\alpha}_1 + \cdots + a_{k,k+1}\boldsymbol{\alpha}_k + a_{k+1,k+1}\boldsymbol{\alpha}_{k+1} + \cdots + a_{n,k+1}\boldsymbol{\alpha}_n \\
&\vdots \\
f(\boldsymbol{\alpha}_n) &= a_{1n}\boldsymbol{\alpha}_1 + \cdots + a_{kn}\boldsymbol{\alpha}_k + a_{k+1,n}\boldsymbol{\alpha}_{k+1} + \cdots + a_{nn}\boldsymbol{\alpha}_n
\end{aligned}$$

所以在基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 下，线性变换 f 的矩阵为

$$\boldsymbol{A} = \left(\begin{array}{ccc|ccc}
\lambda_0 & \cdots & 0 & a_{1,k+1} & \cdots & a_{1n} \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & \lambda_0 & a_{k,k+1} & \cdots & a_{kn} \\
\hline
0 & \cdots & 0 & a_{k+1,k+1} & \cdots & a_{k+1,n} \\
\vdots & \vdots & \vdots & \vdots & \boldsymbol{A}_0 & \vdots \\
0 & \cdots & 0 & a_{n,k+1} & \cdots & a_{nn}
\end{array}\right)$$

由于 f 的特征多项式 $f(\lambda)=|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-\lambda_0)^k|\lambda\boldsymbol{E}-\boldsymbol{A}_0|$，记 $\varphi(\lambda)=|\lambda\boldsymbol{E}-\boldsymbol{A}_0|$，这是一个 $n-k$ 次的多项式，因此

$$f(\lambda)=(\lambda-\lambda_0)^k\varphi(\lambda)$$

所以特征值 λ_0 在多项式 $f(\lambda)$ 中的重数（即 f 的代数重数）不小于 k（即 f 的几何重数）；考虑到 λ_0 无特殊性，去掉足标，有 f 的任意特征值的几何重数不大于它的代数重数.

定理 2.6.3 对于矩阵 $\boldsymbol{A}$，下列结论成立：

(1) 设 $\boldsymbol{A}\in P^{n\times n}$，则 $\boldsymbol{A}$ 与 $\boldsymbol{A}^{\mathrm{T}}$ 有相同的特征多项式和相同的特征值；

(2) 若 λ 为 $\boldsymbol{A}$ 的一个特征值，$\boldsymbol{\alpha}$ 为其所对应的特征向量，则 $k\lambda$ 与 λ^m 分别为 $k\boldsymbol{A}$ 与 $\boldsymbol{A}^m$ 的特征值，$\boldsymbol{\alpha}$ 仍为其所对应的特征向量（k=常数，$m\in\mathbf{Z}^+$）；

(3) 若 λ 为可逆矩阵 $\boldsymbol{A}$ 的某个特征值，$\boldsymbol{\alpha}$ 为其所对应的特征向量，则 $\dfrac{1}{\lambda}$（$\lambda\neq0$）为 $\boldsymbol{A}^{-1}$ 的特征值，$\boldsymbol{\alpha}$ 仍为其所对应的特征向量；

(4) 如果 $\boldsymbol{A}\backsim\boldsymbol{B}$，则 $f(\boldsymbol{A})=f(\boldsymbol{B})$.

在有限维线性空间中不同的基下，同一个线性变换具有不同的矩阵. 但是由于这些矩阵是相似的，因此其特征多项式相同. 换言之，特征多项式 $f(\boldsymbol{A})$ 与基无关，它仅由线性变换 f 决定.

定理 2.6.4 矩阵 $\boldsymbol{A}$ 的相异的特征值所对应的特征向量线性无关.
也就是说，如果 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 是线性变换 f 的属于不同的特征值 λ_1，λ_2，…，λ_n 的特征向量，则向量组 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 线性无关.

证 （对 n 应用数学归纳法）
当 $n=1$ 时，由于特征向量不为零，因此结论成立；
设 $n>1$ 且对 $n-1$ 时结论正确；
对于 n，设 λ_1，λ_2，…，λ_n 是 f 的两两相异的特征值，$\boldsymbol{\alpha}_i$ 是属于 λ_i 的特征向量，即

$$f(\boldsymbol{\alpha}_i)=\lambda_i\boldsymbol{\alpha}_i \quad (i=1,2,\cdots,n)$$

如果

$$a_1\boldsymbol{\alpha}_1+a_2\boldsymbol{\alpha}_2+\cdots+a_n\boldsymbol{\alpha}_n=0 \quad (a_i\in P,i=1,2,\cdots,n) \tag{2.6.6}$$

以 λ_n 乘以式（2.6.6）得

$$a_1\lambda_n\boldsymbol{\alpha}_1+a_2\lambda_n\boldsymbol{\alpha}_2+\cdots+a_n\lambda_n\boldsymbol{\alpha}_n=0 \tag{2.6.7}$$

对式（2.6.6）实施线性变换 f，由 $f(\boldsymbol{\alpha}_i)=\lambda_i\boldsymbol{\alpha}_i$，有

$$a_1\lambda_1\boldsymbol{\alpha}_1+a_2\lambda_2\boldsymbol{\alpha}_2+\cdots+a_n\lambda_n\boldsymbol{\alpha}_n=0 \tag{2.6.8}$$

式（2.6.8）减式（2.6.7）得到

$$a_1(\lambda_1-\lambda_n)\boldsymbol{\alpha}_1+a_2(\lambda_2-\lambda_n)\boldsymbol{\alpha}_2+\cdots+a_{n-1}(\lambda_{n-1}-\lambda_n)\boldsymbol{\alpha}_{n-1}=0$$

因为 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_{n-1}$线性无关，所以 $a_i(\lambda_i-\lambda_n)=0 \quad (i=1,2,\cdots,n-1)$，但 λ_1，λ_2，…，λ_n 两两相异，因此 $a_1=a_2=\cdots=a_{n-1}=0$，代入式（2.6.6），且知 $\boldsymbol{\alpha}_n\neq\boldsymbol{0}$，所以 $a_n=0$，即 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 线性无关.

推论 2.6.1 设线性变换 $f\in L(\boldsymbol{V})$，如果 f 的特征多项式 $f_{\boldsymbol{A}}(x)$在数域 P 内有 n 个单根，则存在着一个 n 阶可逆矩阵 $\boldsymbol{B}$，使得

$$\boldsymbol{B}^{-1}\boldsymbol{AB}=\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{pmatrix}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

式中：$\boldsymbol{A}$ 为 f 在某一个基下的矩阵.

证 设 $f_{\boldsymbol{A}}(x)=(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_n)$，对任意的 λ_i，选取一个 $\boldsymbol{\alpha}_i(i=1,2,\cdots,n)$.

由定理 2.6.4 可知 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 线性无关，并且 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 构成 V_n 的一个基，而 f 关于此基的矩阵为

$$\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{pmatrix}$$

又因为 f 在不同的基下的矩阵之间的相似关系，则对于 f 的任意一个矩阵 $\boldsymbol{A}$，存在着可逆矩阵 $\boldsymbol{B}$，使得

$$\boldsymbol{B}^{-1}\boldsymbol{AB}=\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{pmatrix}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

(请读者注意，这个推论仅为 n 阶矩阵可以对角化的充分条件，而非必要条件. 例如单位矩阵 $\boldsymbol{E}$ 是对角矩阵，但它却只有 n 重特征值 1).

我们都知道，在一般情况下，n 阶矩阵未必有 n 个互异的特征值.

定义 2.6.5 设线性变换 $f\in L(\boldsymbol{V})$，f 有 t 个相异的特征值 λ_1，λ_2，…，λ_t，而 $t\leqslant n$，$\boldsymbol{V}_{\lambda_1}$，$\boldsymbol{V}_{\lambda_2}$，…，$\boldsymbol{V}_{\lambda_t}$ 为相应的特征子空间

$$\dim\boldsymbol{V}_{\lambda_i}=r_i \quad (i=1,2,\cdots,t)$$

其基为 $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\cdots,\boldsymbol{\alpha}_{i_{r_i}}\ (i=1,2,\cdots,t)$，则向量组 $\boldsymbol{\alpha}_{i_1}$，$\boldsymbol{\alpha}_{i_2}$，…，$\boldsymbol{\alpha}_{i_{r_i}}$ $(i=1,2,\cdots,t)$叫做线性变换 f 的一个**特征向量系**. 并且有：

(1) 特征向量系线性无关；

(2) 当特征向量系所含的向量个数为 n 时，称为**完全特征向量系**.

显然，如果 f 有完全特征向量系，以此向量系为基，则 f 在此基下的矩阵 $\boldsymbol{A}$ 是对角形矩阵，此时称 f 的矩阵 $\boldsymbol{A}$ 为**单纯矩阵**，并且 $\boldsymbol{A}$ 的所有特征值的几何重数与代数重数相等；如果 f 的特征向量系中向量的个数小于线性空间的维数 n，则 f 在任意一个基下的矩阵 $\boldsymbol{A}$ 都不能成为对角形矩阵，此时称 f 的矩阵 $\boldsymbol{A}$ 为**非单纯矩阵**（或**亏损矩阵**）.

定理 2.6.5 线性变换 f 的矩阵在某一个基下是对角形矩阵的充分且必要条件为 f 有完全特征向量系. 即 $\sum\limits_{i=1}^{t}\dim\boldsymbol{V}_{\lambda_i}=\dim\boldsymbol{V}=n.$

定理 2.6.6 矩阵 $\boldsymbol{A}$ 可以对角化的充分必要条件是 $\boldsymbol{A}$ 的所有特征值的几何重数与代数重数相等.

由前面的讨论可以看出，将一个矩阵化为对角形矩阵是不太容易做到的. 但是，将一个矩阵化为三角形矩阵是较为容易的，我们将在本书的第四章中详细讨论这一问题（当然这些讨论是对 n 阶方阵进行的).

特征值的应用十分广泛，但其计算却不是一件容易的事. 下面我们介绍一种利用矩阵本身的元素来确定其特征值的较为精确的近似值及其分布区域的方法，即所谓的圆盘定理.

定理 2.6.7 （圆盘定理 1）设矩阵 $\boldsymbol{A}=(a_{ij})\in\mathbf{C}^n$，并且 $R_i=\sum\limits_{\substack{j=1\\j\neq i}}^{n}|a_{ij}|=\sum\limits_{j=1}^{n}|a_{ij}|-|a_{ii}|\quad(i=1,2,\cdots,n)$，将复平面上中心在 a_{ii}，半径为 R_i 的圆形区域记为 S_i，即

$$S_i=\{z\,|\,|z-a_{ii}|\leqslant R_i\}\ (i=1,2,\cdots,n)$$

矩阵 $\boldsymbol{A}$ 的特征值为 λ_1，λ_2，…，λ_n，则 $\lambda_i\in S=S_1\cup S_2\cup\cdots\cup S_n\quad(i=1,2,\cdots,n)$，$S$ 叫做矩阵 $\boldsymbol{A}$ 的（行）Gerschgorin 区，$S_i=\{z\,|\,|z-a_{ii}|\leqslant R_i\}\quad(i=1,2,\cdots,n)$ 叫做矩阵 $\boldsymbol{A}$ 的（行）Gerschgorin 圆盘.

证 设 λ 为矩阵 $\boldsymbol{A}$ 的任一特征值，则 $\boldsymbol{Ax}=\lambda\boldsymbol{x}$，$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，则 $\sum\limits_{j=1}^{n}a_{ij}x_j=\lambda x_j\quad(i=1,2,\cdots,n)$，因此 $(\lambda-a_{ii})x_j=\sum\limits_{\substack{j=1\\j\neq i}}^{n}a_{ij}x_j.$

令 $|x_k|=\max\{|x_1|,|x_2|,\cdots,|x_n|\}$，因为 $\boldsymbol{x}\neq 0$，所以 $|x_k|\neq 0$，由 $(\lambda-a_{kk})x_k=\sum\limits_{\substack{j=1\\j\neq k}}^{n}a_{kj}\boldsymbol{x}_j$，有 $|\lambda-a_{kk}|=\left|\dfrac{1}{x_k}\sum\limits_{\substack{j=1\\j\neq k}}^{n}a_{kj}x_j\right|\leqslant\sum\limits_{\substack{j=1\\j\neq k}}^{n}|a_{kj}|\dfrac{|x_j|}{|x_k|}\leqslant\sum\limits_{\substack{j=1\\j\neq k}}^{n}|a_{kj}|=R_k$，

所以 $\lambda\in S_k\subset S$，即 $\lambda\in S.$

【例 2.6.6】 试估计矩阵 $\boldsymbol{A}=\begin{pmatrix}1&-\frac{1}{2}&-\frac{1}{2}&0\\-\frac{1}{2}&\frac{3}{2}&i&0\\0&-\frac{i}{2}&5&\frac{i}{2}\\-1&0&0&5i\end{pmatrix}$ 的特征值的分布范围.

解 $\because a_{11}=1$，$a_{22}=\dfrac{3}{2}$，$a_{33}=5$，$a_{44}=5i$，

$\therefore S_1$：$|z-1|\leqslant|a_{12}|+|a_{13}|+|a_{14}|=\frac{1}{2}+\frac{1}{2}+0=1$；

S_2：$\left|z-\frac{3}{2}\right|\leqslant|a_{21}|+|a_{23}|+|a_{24}|=\frac{1}{2}+|i|+0=\frac{3}{2}$；

S_3：$|z-5|\leqslant|a_{31}|+|a_{32}|+|a_{34}|=0+\left|-\frac{i}{2}\right|+\left|\frac{i}{2}\right|=1$；

S_4：$|z-5i|\leqslant|a_{41}|+|a_{42}|+|a_{43}|=|-1|+0+0=1$.

由此可知矩阵 $\boldsymbol{A}$ 的四个特征值分别分布在以上四个圆盘中．此例中，圆盘 S_1 与 S_2 的交集合不空，S_1 与 S_2 的并集合构成一个连通区域．我们也称由相交的几个 Gerschgorin 圆盘所构成的连通区域为 S 的一个**连通部分**，孤立的 Gerschgorin 圆盘也视为一个连通部分.

一般地，由 k 个 Gerschgorin 圆盘所构成的连通部分内，必有 k 个特征值.

定理 2.6.8 （圆盘定理 2）设 G 是由矩阵 $\boldsymbol{A}$ 的 m 个 Gerschgorin 圆盘所构成的任一连通部分，那么 G 中有且仅有矩阵 $\boldsymbol{A}$ 的 m 个特征值.

证 （略）.

由于矩阵 $\boldsymbol{A}$ 与其转置矩阵 $\boldsymbol{A}^{\mathrm{T}}$ 有相同的特征值，因此可以得到关于矩阵 $\boldsymbol{A}$ 的列的圆盘定理.

设矩阵 $\boldsymbol{A}$ 的列元素之和为 Gerschgorin 圆盘的半径，即 $\rho_i=\sum\limits_{\substack{i=1\\ i\neq j}}^{n}|a_{ij}|\quad(j=1,2,\cdots,n)$，则有以下推论.

推论 2.6.2 矩阵 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$ 的所有特征值含于 n 个列 Gerschgorin 圆盘 $G_i=\{z\,|\,|z-a_{ii}|\leqslant\rho_i\}$ 的并集合 $G=G_1\cup G_2\cup\cdots\cup G_n$ 中，G 叫做矩阵 $\boldsymbol{A}$ 的（列）Gerschgorin 区，$G_i=\{z\,|\,|z-a_{ii}|\leqslant\rho_i\}\quad(i=1,2,\cdots,n)$ 叫做矩阵 $\boldsymbol{A}$ 的（列）Gerschgorin 圆盘.

如果有 s 个圆盘构成一个孤立于其他圆盘的连通部分，则矩阵 $\boldsymbol{A}$ 有且仅有 s 个特征值在这个连通部分中.

推论 2.6.3 矩阵 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$ 的全部特征值含在复平面的区域 $T=\left(\bigcup\limits_{i=1}^{n}S_i\right)\cap\left(\bigcup\limits_{j=1}^{n}G_j\right)$ 之中.

【例 2.6.7】 试计算矩阵 $\boldsymbol{A}=\begin{pmatrix}1 & 0.1 & 0.2 & 0.3\\ 0.5 & 3 & 0.1 & 0.2\\ 1 & 0.3 & -1 & 0.5\\ 0.2 & -0.3 & -0.1 & -4\end{pmatrix}$ 的特征值的分布范围.

解 $\boldsymbol{A}$ 的行 Gerschgorin 圆盘为 (1，0.6)，(3，0.8)，(−1，1.8)，(−4，0.6)，$\boldsymbol{A}$ 的列 Gerschgorin 圆盘为 (1，1.7)，(3，0.7)，(−1，0.4)，(−4，1)，所以 $T_1=(1,\ 0.6)$，$T_2=(3,\ 0.7)$，$T_3=(-1,\ 0.4)$，$T_4=(-4,\ 0.6)$.

由于 $T_i\cap T_j=\boldsymbol{\Phi}\quad(i\neq j)$，所以每个 T_i 中有且仅有矩阵 $\boldsymbol{A}$ 的一个特征值.

在本节的最后，我们将 $n\times n$ 矩阵 $\boldsymbol{A}$ 的特征值的性质汇总如下：

(1) 矩阵 $\boldsymbol{A}$ 共有 n 个特征值，其中相同特征值按重数计算；

(2) 若 $\boldsymbol{A}$ 是实对称矩阵或是 Hermitian 矩阵，则其全部特征值均为实数；

(3) 若 $\boldsymbol{A}=\mathrm{diag}(a_{11},a_{22},\cdots)$，则其特征值即为 a_{11}，a_{22}，…；

(4) 若 $\boldsymbol{A}$ 为三角矩阵，则其对角元素即为全部特征值；

(5) 若 λ 是 $\boldsymbol{A}$ 的特征值，则 λ 是 $\boldsymbol{A}^{\mathrm{T}}$ 的特征值；

(6) 若 λ 是 $\boldsymbol{A}$ 的特征值，则 λ 是 $\boldsymbol{A}^{\mathrm{H}}$ 的特征值；

(7) 若 λ 是 $\boldsymbol{A}$ 的特征值，则 $\lambda+\sigma^2$ 是 $\boldsymbol{A}+\sigma^2\boldsymbol{E}$ 的特征值；

(8) 若 λ 是 $\boldsymbol{A}$ 的特征值，则 $1/\lambda$ 也是 $\boldsymbol{A}^{-1}$ 的特征值；

(9) 幂等矩阵 $\boldsymbol{A}^2$ 的所有特征值取 0 或 1；

(10) 若 $\boldsymbol{A}$ 是实正交矩阵，则其所有特征值位于单位圆上；

(11) 若 $\boldsymbol{A}$ 奇异，则它至少有一个特征值是 0；

(12) 若 $\boldsymbol{A}$ 非奇异，则其全部特征值均非零；

(13) $\boldsymbol{A}$ 的特征值之和等于 $\boldsymbol{A}$ 的迹；

(14) 与互不相同的特征值对应的非零特征向量线性无关；

(15) 若 $\boldsymbol{A}$ 有 r 个非零特征值，则 $\mathrm{rank}\boldsymbol{A}\geqslant r$；

(16) 若 0 是 $\boldsymbol{A}$ 的特征值，并且重数为 1，则 $\mathrm{rank}\boldsymbol{A}=n-1$；

(17) 若 $\mathrm{rank}(\boldsymbol{A}-\lambda\boldsymbol{E})\leqslant n-1$，则 λ 是 $\boldsymbol{A}$ 的特征值；

(18) $\boldsymbol{A}$ 的全部特征值之积等于该矩阵的行列式之值；

(19) Hermitian 矩阵 $\boldsymbol{A}$ 是正定（或半正定），当且仅当其特征值是正（或非负）的；

(20) 若 $\boldsymbol{A}$ 的特征值互不相同，则存在一个相似矩阵 $\boldsymbol{S}^{-1}\boldsymbol{AS}=\boldsymbol{D}$（对角阵），并且其对角元素就是 $\boldsymbol{A}$ 的全部特征值；

(21) 若 λ_1，λ_2，…，λ_n 是 $\boldsymbol{A}$ 的全部特征值，则 $\prod_{i=1}^{n}(\boldsymbol{A}-\lambda_i\boldsymbol{E})=0$（这个结论就是后面将要讲到的 Cayley-Hamilton 定理）；

(22) $\boldsymbol{A}$ 的任一特征值的几何重数都不大于其代数重数；

(23) 若 λ 是 $\boldsymbol{A}$ 的一个特征值，并且 $\boldsymbol{B}$ 非奇异，则 λ 也是 $\boldsymbol{B}^{-1}\boldsymbol{AB}$ 的一个特征值，但是对应的特征向量一般不同；

(24) 若 λ 是 $\boldsymbol{A}$ 的一个特征值，并且 $\boldsymbol{B}$ 是酉矩阵，则 λ 也是 $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{AB}$ 的一个特征值，但是对应的特征向量一般不同；

(25) 若 λ 是 $\boldsymbol{A}$ 的一个特征值，并且 $\boldsymbol{B}$ 是正交矩阵，则 λ 也是 $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{AB}$ 的一个特征值，但是对应的特征向量一般不同；

(26) $\boldsymbol{A}$ 的最大特征值以该矩阵的列元素之和的最大值为界；

(27) 若 $|\lambda_i|<1, i=1,2,\cdots,n$，则 $\boldsymbol{A}\pm\boldsymbol{E}$ 非奇异；

(28) 若 λ 是 $\boldsymbol{AB}$ 的特征值（其中 $\boldsymbol{A}$ 是 $m\times n$ 矩阵，$n\geqslant m$；$\boldsymbol{B}$ 是 $n\times m$ 矩阵），则 λ 也是 $\boldsymbol{BA}$ 的特征值；

(29) 若 $\lambda\neq0$ 是 $\boldsymbol{BA}$ 的特征值（其中 $\boldsymbol{A}$ 是 $m\times n$ 矩阵，$n\geqslant m$；$\boldsymbol{B}$ 是 $n\times m$ 矩阵），则 λ 也是 $\boldsymbol{AB}$ 的特征值；

(30) 若 λ_1，λ_2，…，λ_m 是 $\boldsymbol{AB}$ 的特征值，则 $\boldsymbol{BA}$ 的 n 个特征值为 λ_1，λ_2，…，λ_m，0，…，0；

(31) 若 λ 是 $\boldsymbol{A}$ 的特征值，则矩阵多项式 $f(\boldsymbol{A})=\boldsymbol{A}^n+c_1\boldsymbol{A}^{n-1}+\cdots+c_{n-1}\boldsymbol{A}+c_n\boldsymbol{E}$ 的特征值为 $f(\lambda)=\lambda^n+c_1\lambda^{n-1}+\cdots+c_{n-1}\lambda+c_n$.

习 题 2

2.1 判别下列变换中，哪些是线性变换，哪些不是线性变换：

(1) $\begin{aligned} T: \quad & \mathbf{C} \to \mathbf{C} \\ & z \mapsto \bar{z} \end{aligned}$；

(2) $\begin{aligned} T: \quad & \boldsymbol{P}_n[x] \to \boldsymbol{P}_n[x] \\ & p(x) \mapsto p(x+a) \end{aligned}$，其中 a 是固定实数；

(3) $\begin{aligned} T: \quad & \mathbf{C}^{m\times n} \to \mathbf{C}^{m\times n} \\ & X \mapsto \boldsymbol{P}X \end{aligned}$，其中 $\boldsymbol{P}$ 是固定 $m\times m$ 阶矩阵；

(4) $\begin{aligned} T: \quad & \mathbf{C}[a,b] \to \mathbf{C}[a,b] \\ & f(x) \mapsto \int_a^x K(x,y)f(y)dy \end{aligned}$，其中 $K(x,y)$ 是一固定的连续函数；

(5) $\begin{aligned} T: \quad & \boldsymbol{N} \to \boldsymbol{N} \\ & n \mapsto 2n \end{aligned}$.

2.2 设 $\boldsymbol{T}$ 是 $\boldsymbol{R}^3$ 的线性变换，$\boldsymbol{T}(x_1,x_2,x_3)=(x_1+2x_2-x_3,x_2+x_3,x_1+x_2-2x_3)$. 求：

(1) $\boldsymbol{R}(\boldsymbol{T})$ 的一个基和维数；

(2) $\boldsymbol{N}(\boldsymbol{T})$ 的一个基和维数 .

2.3 求线性变换

$$\begin{aligned} T: \quad & \mathbf{R}^{2\times 2} \to \mathbf{R}^{2\times 2} \\ & \boldsymbol{X} \mapsto \boldsymbol{PX} \end{aligned}, \boldsymbol{P} = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix}$$

在基 $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$，$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$，$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$，$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ 下的矩阵 .

2.4 设线性变换 $\boldsymbol{T}$ 在基 $\boldsymbol{e}_1$，$\boldsymbol{e}_2$，$\boldsymbol{e}_3$ 下的矩阵为

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

求 $\boldsymbol{T}$ 在基 $\boldsymbol{e}_1$，$2\boldsymbol{e}_2$，$\boldsymbol{e}_3+2\boldsymbol{e}_1$ 下的矩阵 .

2.5 $\mathbf{R}^3$ 中，线性变换 σ 关于基 $\boldsymbol{\alpha}_1=(-1,1,1)$，$\boldsymbol{\alpha}_2=(1,0,-1)$，$\boldsymbol{\alpha}_3=(0,1,1)$ 的矩阵为

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 2 & 1 \end{pmatrix}$$

求：

(1) σ 关于标准基 $\boldsymbol{\varepsilon}_1$，$\boldsymbol{\varepsilon}_2$，$\boldsymbol{\varepsilon}_3$ 的矩阵；

(2) 设 $\boldsymbol{\alpha}=\boldsymbol{\alpha}_1+6\boldsymbol{\alpha}_2-\boldsymbol{\alpha}_3$，$\boldsymbol{\beta}=\boldsymbol{\varepsilon}_1-\boldsymbol{\varepsilon}_2+\boldsymbol{\varepsilon}_3$，求 $\sigma(\boldsymbol{\alpha})$，$\sigma(\boldsymbol{\beta})$ 关于基 $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3\}$ 的坐标 .

2.6 二维线性空间 $\mathbf{R}^2$ 中线性变换 $\boldsymbol{A}$ 在基 $\boldsymbol{\alpha}_1=(1,2)$，$\boldsymbol{\alpha}_2=(2,1)$ 下的矩阵是 $\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$，线性变换 $\boldsymbol{B}$ 在基 $\boldsymbol{\beta}_1=(1,1)$，$\boldsymbol{\beta}_2=(1,2)$ 下的矩阵是 $\begin{pmatrix} 3 & 3 \\ 2 & 4 \end{pmatrix}$，求 $\boldsymbol{A}+\boldsymbol{B}$ 在基 $\boldsymbol{\beta}_1=(1,1)$，$\boldsymbol{\beta}_2=(1,2)$ 下的矩阵.

2.7 设

$$T:\quad \begin{aligned} P_n[x] &\to P_{n+1}[x] \\ p(x) &\mapsto \int_a^x p(x-a)\mathrm{d}x \end{aligned}$$

其中 a 是固定实数．

(1) 在 V 中取一个基，写出线性算子 T 在该基下的矩阵；

(2) 线性算子 T 的值域和核，以及它们的维数．

2.8 设

$$T:\quad \begin{aligned} \mathbf{R}^{2\times 2} &\to \mathbf{R}^{2\times 2} \\ X &\mapsto PX \end{aligned}, P=\begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix}$$

求：

(1) 在 $\mathbf{R}^{2\times 2}$ 中取一组基，写出线性变换 T 在该基下的矩阵；

(2) 线性变换 T 的值域和核，以及它们的维数；

(3) 线性变换 T 特征值和特征向量；

(4) 线性变换 T 的不变子空间．

2.9 设 T 是 V 上的线性变换，e_1，e_2，…，e_n 是 V 的一组基，证明：

(1) T 可逆的充分必要条件是 Te_1，Te_2，…，Te_n 线性无关；

(2) T 可逆的充分必要条件是 T 的核空间为零空间；

(3) T 可逆的充分必要条件是 T 没有零特征值．

2.10 设 T 是 V 上的线性变换，$\alpha\in V$. 如果 $T^n\alpha\neq 0$，$T^{n+1}\alpha=0$，证明：α，$T\alpha$，…，$T^n\alpha$ 线性无关．

2.11 设 $A\in L$ (V)，$0\neq\alpha\in V$，证明：如果 α，$T\alpha$，…，$T^{m-1}\alpha$ 线性无关，而 α，$T\alpha$，…，$T^m\alpha$ 线性相关，那么证明：

(1) $\mathrm{span}\{\alpha, T\alpha, \cdots, T^m\alpha\}$ 是 T-子空间；

(2) $\mathrm{span}\{\alpha, T\alpha, \cdots, T^{m-1}\alpha\}$ 是包含 α 的最小 T-子空间．

2.12 设 $A=\begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$ 是非奇异的实矩阵．证明：$R^n=\ker(A_1)\oplus\ker(A_2)$．

2.13 用圆盘定理估计矩阵

$$\begin{pmatrix} 3 & 0.4 & 0.2 & 0.2 \\ 0.2 & 3 & 0.1 & 0.1 \\ 0.4 & 0.5 & 4 & 0.5 \\ -0.3 & 1 & 0.5 & 1 \end{pmatrix}$$

的特征值的分布范围．

2.14 用圆盘定理证明如下矩阵是可以对角化的非奇异矩阵.

$$\begin{pmatrix} 1 & 0.4 & 0.1 & 0.2 \\ 0.2 & 3 & 0.1 & 0.1 \\ 0.4 & 0.5 & 6 & 0.5 \\ -0.3 & 1 & 0.5 & 12 \end{pmatrix}$$

第三章 Jordan 标准形

在第二章的结尾我们讨论了如何将单纯矩阵化为对角形矩阵，于是对角形矩阵就可以作为与单纯矩阵相似的一类矩阵的标准形．同时我们也知道，非单纯矩阵是不能化为对角形矩阵的，本章将使用λ-矩阵来研究非单纯矩阵的化简，并希望找出与非单纯矩阵相似的一类矩阵的标准形．为此首先讨论所谓的λ-矩阵及其相关的一些性质．

§3.1 λ-矩阵

定义 3.1.1 以λ的复系数多项式为元素的矩阵，叫做λ-矩阵．记作 $\boldsymbol{A}(\lambda),\boldsymbol{B}(\lambda)$，….

例如，$\begin{pmatrix}\lambda^2-3\lambda+4 & 4\lambda+5\\ 2\lambda-1 & 6\lambda^3-7\end{pmatrix}$是λ-矩阵，而$\begin{pmatrix}\sin\lambda & 4\lambda+5\\ 2\lambda-1 & 7\end{pmatrix}$不是λ-矩阵．又如 $\lambda\boldsymbol{E}-\boldsymbol{A}$ 就是一个λ-矩阵．

此外，一般的数字矩阵也可以视为λ-矩阵，只不过它的每一个元素都是λ的零次多项式．

定义 3.1.2 如果λ-矩阵 $\boldsymbol{A}(\lambda)$ 中有一个 r ($r\geqslant 1$) 阶子式不为 0，而所有 $r+1$ 阶子式（如果有的话）全为 0，则称 $\boldsymbol{A}(\lambda)$ 的秩为 r. 规定零矩阵的秩为 0.

定义 3.1.3 对λ-矩阵 $\boldsymbol{A}(\lambda)$，如果存在λ-矩阵 $\boldsymbol{B}(\lambda)$，使得

$$\boldsymbol{A}(\lambda)\boldsymbol{B}(\lambda)=\boldsymbol{B}(\lambda)\boldsymbol{A}(\lambda)=\boldsymbol{E} \tag{3.1.1}$$

则称 $\boldsymbol{A}(\lambda)$ 是可逆的，并称 $\boldsymbol{B}(\lambda)$ 是 $\boldsymbol{A}(\lambda)$ 的逆矩阵，记为 $\boldsymbol{A}^{-1}(\lambda)$．

在有些书中，将可逆的λ-矩阵，称为单位模阵．

定理 3.1.1 一个 n 阶λ-矩阵 $\boldsymbol{A}(\lambda)$ 是可逆的充分且必要条件为行列式 $|\boldsymbol{A}(\lambda)|$ 是一个非零的数．

证 充分性 设 $d=|\boldsymbol{A}(\lambda)|$ 是一个非零的数．$\boldsymbol{A}(\lambda)$ 的伴随矩阵 $\mathrm{adj}(\boldsymbol{A}(\lambda))$ 也是一个λ-矩阵，且

$$\boldsymbol{A}(\lambda)\cdot\frac{1}{d}\mathrm{adj}(\boldsymbol{A}(\lambda))=\frac{1}{d}\mathrm{adj}(\boldsymbol{A}(\lambda))\cdot\boldsymbol{A}(\lambda)=\boldsymbol{E}$$

因此，$\boldsymbol{A}(\lambda)$ 可逆，且 $(\boldsymbol{A}(\lambda))^{-1}=\mathrm{adj}(\boldsymbol{A}(\lambda))/d$．

必要性 若 $\boldsymbol{A}(\lambda)$ 可逆，在 (3.1.1) 的两边取行列式，则 $|\boldsymbol{A}(\lambda)||\boldsymbol{B}(\lambda)|=|\boldsymbol{E}|=1$. 因为 $|\boldsymbol{A}(\lambda)|$ 与 $|\boldsymbol{B}(\lambda)|$ 都是λ的多项式，所以由它们的乘积是 1 推知，它们都是零次多项式，即为非零的数．

定义 3.1.4 以下三类变换称为λ-矩阵的初等变换：

(1) 互换两行（列）；

(2) 某行（列）乘非零常数；

(3) 某行（列）乘多项式后加到另一行（列）．

定义 3.1.5 如果λ-矩阵 $\boldsymbol{A}(\lambda)$ 经过有限次初等变换后可变成 $\boldsymbol{B}(\lambda)$，则称 $\boldsymbol{B}(\lambda)$ 与 $\boldsymbol{A}(\lambda)$ 等价．$\boldsymbol{B}(\lambda)$ 与 $\boldsymbol{A}(\lambda)$ 等价，则记 $\boldsymbol{B}(\lambda)\sim\boldsymbol{A}(\lambda)$．

λ-矩阵的等价关系与一般的等价关系一样，满足自反律、对称律、传递律．

定义 3.1.6 单位阵经过一次初等变换后所得到的矩阵称为初等矩阵．初等矩阵是可逆矩阵．

利用初等变换与初等矩阵的关系，可得下面定理．

定理 3.1.2 设 $\boldsymbol{A}(\lambda),\boldsymbol{B}(\lambda)$ 是 $m\times n$ 阶 λ-矩阵，则 $\boldsymbol{A}(\lambda)\sim\boldsymbol{B}(\lambda)$ 的充分且必要条件是有一个 m 阶可逆 λ-矩阵 $\boldsymbol{P}(\lambda)$ 与一个 n 阶可逆 λ-矩阵 $\boldsymbol{Q}(\lambda)$，使得 $\boldsymbol{B}(\lambda)=\boldsymbol{P}(\lambda)\boldsymbol{A}(\lambda)\boldsymbol{Q}(\lambda)$．

定理 3.1.3 如果 λ-矩阵 $\boldsymbol{A}(\lambda)$ 与 $\boldsymbol{B}(\lambda)$ 等价，则 $\operatorname{rank}\boldsymbol{A}(\lambda)=\operatorname{rank}\boldsymbol{B}(\lambda)$，但反之不然．

如设 $\boldsymbol{A}(\lambda)=\begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix},\boldsymbol{B}(\lambda)=\begin{pmatrix}1 & -\lambda\\ 1 & \lambda\end{pmatrix}$，因为 $\det\boldsymbol{A}(\lambda)\neq 0,\det\boldsymbol{B}(\lambda)\neq 0$；并且 $\operatorname{rank}\boldsymbol{A}(\lambda)=\operatorname{rank}\boldsymbol{B}(\lambda)=2$．由矩阵的初等变换可知，如果 $\boldsymbol{A}(\lambda)$ 与 $\boldsymbol{B}(\lambda)$ 等价，则 $\det\boldsymbol{A}(\lambda)$ 与 $\det\boldsymbol{B}(\lambda)$ 之间只能相差一个不为零的常数因子，而 $\boldsymbol{A}(\lambda)$ 与 $\boldsymbol{B}(\lambda)$ 不满足这一条件，所以 $\boldsymbol{A}(\lambda)$ 与 $\boldsymbol{B}(\lambda)$ 不等价．

我们现在来讨论如何将 λ-矩阵化为标准形．

引理 3.1.1 如果 λ-矩阵 $\boldsymbol{A}(\lambda)$ 中的元素 $a_{11}(\lambda)\neq 0$，并且 $\boldsymbol{A}(\lambda)$ 中至少有一个元素不能被其整除，则必存在一个与 $\boldsymbol{A}(\lambda)$ 等价的 λ-矩阵 $\boldsymbol{B}(\lambda)$，并且 $\boldsymbol{B}(\lambda)$ 中的元素 $b_{11}(\lambda)\neq 0$，同时多项式 $b_{11}(\lambda)$ 的次数小于 $a_{11}(\lambda)$ 的次数．

证 不妨设 $\boldsymbol{A}(\lambda)$ 中第 1 行上有元素 $a_{1j}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除，则 $a_{1j}(\lambda)=a_{11}\varphi(\lambda)+b_{11}(\lambda)$，$b_{11}(\lambda)\neq 0$ 并且 $b_{11}(\lambda)$ 的次数小于 $a_{11}(\lambda)$ 的次数；将 $\boldsymbol{A}(\lambda)$ 中第 1 列的元素乘以 $-\varphi(\lambda)$ 后，加到第 j 列上，并将第 1 列与第 j 列互换，得到

$$\boldsymbol{B}(\lambda)=\begin{pmatrix}b_{11}(\lambda) & * & \cdots & *\\ * & * & \cdots & *\\ \vdots & \vdots & \vdots & \vdots\\ * & * & \cdots & *\end{pmatrix}$$

并且 $\boldsymbol{B}(\lambda)$ 与 $\boldsymbol{A}(\lambda)$ 等价，在 $\boldsymbol{B}(\lambda)$ 中 $b_{11}(\lambda)\neq 0,b_{11}(\lambda)$ 的次数小于 $a_{11}(\lambda)$ 的次数；同理，若 $\boldsymbol{A}(\lambda)$ 中第 1 列上有元素 $a_{i1}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除，亦可得到 $\boldsymbol{B}_1(\lambda)$ 满足引理．如果 $\boldsymbol{A}(\lambda)$ 中第 1 行或第 1 列上的所有元素 $a_{i1}(\lambda)$ 都能被 $a_{11}(\lambda)$ 整除，根据定理的条件可知，至少有一个元素 $a_{ij}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除，并且 $i,j\neq 1$. 则有 $a_{i1}(\lambda)=a_{11}(\lambda)\varphi(\lambda)$，将 $\boldsymbol{A}(\lambda)$ 中第 1 行乘以 $-\varphi(\lambda)$ 后，加到第 i 行上，得到 $\boldsymbol{A}'(\lambda)$，其中 $a'_{i1}(\lambda)=0$，$a'_{ij}(\lambda)=a_{ij}(\lambda)-a_{1j}(\lambda)\varphi(\lambda)$，所以 $a_{11}(\lambda)$ 不能整除 $a'_{ij}(\lambda)$；再将第 i 行加到第 1 行上，此时 $a_{11}(\lambda)$ 没有变化，而第 1 行、第 j 列上的元素变为 $[1-\varphi(\lambda)]a_{1j}(\lambda)+a_{ij}(\lambda)$ 也不能被 $a_{11}(\lambda)$ 整除，这就回到了 $\boldsymbol{A}(\lambda)$ 中第 1 行上有元素 $a_{1j}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除的情形．

定理 3.1.4 如果 λ-矩阵 $\boldsymbol{A}(\lambda)\in P^{m\times n}$，并且 $\operatorname{rank}\boldsymbol{A}(\lambda)=r$，那么 $\boldsymbol{A}(\lambda)$ 与矩阵

$$\boldsymbol{B}(\lambda)=\begin{pmatrix}d_1(\lambda) & 0 & \cdots & 0 & 0 & \cdots & 0\\ 0 & d_2(\lambda) & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & d_r(\lambda) & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0\end{pmatrix}$$

等价．其中 $d_i(\lambda)$ 是首项系数为 1 的多项式，并且 $d_i(\lambda)$ 能够整除 $d_{i+1}(\lambda)(i=1,2,\cdots,r-1)$；我们称矩阵 $\boldsymbol{B}(\lambda)$ 是 λ-矩阵 $\boldsymbol{A}(\lambda)$ 的 **Smith 标准形**．

证 如果 $\operatorname{rank}\boldsymbol{A}(\lambda)=0$，则 $\boldsymbol{A}(\lambda)$ 为零矩阵，结论成立．

设 $\operatorname{rank}\boldsymbol{A}(\lambda)=r>0$，并且 $a_{11}(\lambda)\neq 0$（如若不然，可以将矩阵 $\boldsymbol{A}(\lambda)$ 适当调换行或列，使得调换后的矩阵中左上角的元素不为零，并以此元素作为新的 $a_{11}(\lambda)$），如果矩阵 $\boldsymbol{A}(\lambda)$ 中的元素不是全部能被 $a_{11}(\lambda)$ 整除，则由引理 3.1.1，必然存在着一个矩阵 $\boldsymbol{B}^{(1)}(\lambda)\sim\boldsymbol{A}(\lambda)$，并且多项式 $b_{11}^{(1)}(\lambda)$ 的次数小于 $a_{11}(\lambda)$ 的次数．

如果 $b_{11}^{(1)}(\lambda)$ 仍然不能整除 $\boldsymbol{B}^{(1)}(\lambda)$ 中的所有元素，可再次应用引理 3.1.1，逐步降低矩阵 $\boldsymbol{B}^{(i)}(\lambda)$ 中左上角元素 $b_{11}^{(i)}(\lambda)$ 的次数$(i=1,2,\cdots)$，直到求得一个矩阵 $\boldsymbol{B}^{(i)}(\lambda)\sim\boldsymbol{A}(\lambda)$．令矩阵 $\boldsymbol{B}(\lambda)=\boldsymbol{B}^{(i)}(\lambda)$，则多项式 $b_{11}(\lambda)$ 可以整除 $\boldsymbol{B}(\lambda)$ 中的所有元素，并且 $b_{11}(\lambda)$ 是一个首项系数为 1 的多项式．然后将 $\boldsymbol{B}(\lambda)$ 中的第一行元素乘以适当的多项式加到其他各行上，使得第一列的元素中除 $b_{11}(\lambda)$ 外，其余的元素均为零．再将 $\boldsymbol{B}(\lambda)$ 中的第一行的元素（除 $b_{11}(\lambda)\neq 0$ 外），消为零．则得到一个与矩阵 $\boldsymbol{A}(\lambda)$ 等价的矩阵

$$\begin{pmatrix} d_1(\lambda) & 0 & 0 & \cdots & 0 \\ 0 & c_{22}(\lambda) & c_{23}(\lambda) & \cdots & c_{2n}(\lambda) \\ 0 & c_{32}(\lambda) & c_{33}(\lambda) & \cdots & c_{3n}(\lambda) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & c_{m2}(\lambda) & c_{m3}(\lambda) & \cdots & c_{mn}(\lambda) \end{pmatrix}$$

其中 $d_1(\lambda)=b_{11}(\lambda)$ 可以整除 $c_{ij}(\lambda)$，$(i=2,3,\cdots,m;\ j=2,3,\cdots,n)$；

同理，再对子矩阵 $\begin{pmatrix} c_{22}(\lambda) & c_{23}(\lambda) & \cdots & c_{2n}(\lambda) \\ c_{32}(\lambda) & c_{33}(\lambda) & \cdots & c_{3n}(\lambda) \\ \vdots & \vdots & \ddots & \vdots \\ c_{m2}(\lambda) & c_{m3}(\lambda) & \cdots & c_{mn}(\lambda) \end{pmatrix}$ 做类似的变换，得到矩阵

$$\begin{pmatrix} d_1(\lambda) & 0 & 0 & \cdots & 0 \\ 0 & d_2(\lambda) & 0 & \cdots & 0 \\ 0 & 0 & f_{33}(\lambda) & \cdots & f_{3n}(\lambda) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & f_{m3}(\lambda) & \cdots & f_{mn}(\lambda) \end{pmatrix}$$

其中 $d_1(\lambda)$ 可以整除 $d_2(\lambda)$，$d_2(\lambda)$ 可以整除 $f_{ij}(\lambda)$，$(i=3,4,\cdots,m;\ j=3,4,\cdots,n)$；以此类推，最后可以得到矩阵

$$\begin{pmatrix} d_1(\lambda) & & & & & \\ & \ddots & & & & \\ & & d_r(\lambda) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

其中 $d_i(\lambda)$ 可以整除 $d_{i+1}(\lambda)$ $(i=1,2,\cdots,r-1)$．

【例 3.1.1】 用初等变换化 λ-矩阵

$$A(\lambda)=\begin{pmatrix}1-\lambda & 2\lambda-1 & \lambda\\ \lambda & \lambda^2 & -\lambda\\ 1+\lambda^2 & \lambda^3+\lambda-1 & -\lambda^2\end{pmatrix}$$

为 Smith 标准形.

解

$$A(\lambda)\sim\begin{pmatrix}1 & 2\lambda-1 & 1-\lambda\\ 0 & \lambda^2 & \lambda\\ 1 & \lambda^3+\lambda-1 & 1+\lambda^2\end{pmatrix}\sim\begin{pmatrix}1 & 2\lambda-1 & 1-\lambda\\ 0 & \lambda^2 & \lambda\\ 0 & \lambda^3-\lambda & \lambda^2+\lambda\end{pmatrix}\sim\begin{pmatrix}1 & 0 & 0\\ 0 & \lambda^2 & \lambda\\ 0 & \lambda^3-\lambda & \lambda^2+\lambda\end{pmatrix}$$

$$\sim\begin{pmatrix}1 & 0 & 0\\ 0 & \lambda & \lambda^2\\ 0 & \lambda^2+\lambda & \lambda^3-\lambda\end{pmatrix}\sim\begin{pmatrix}1 & 0 & 0\\ 0 & \lambda & 0\\ 0 & \lambda^2+\lambda & -\lambda^2-\lambda\end{pmatrix}\sim\begin{pmatrix}1 & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda^2+\lambda\end{pmatrix}=B(\lambda)$$

【例 3.1.2】 求矩阵 $A=\begin{pmatrix}1 & 2 & 2\\ 2 & 1 & 2\\ 2 & 2 & 1\end{pmatrix}$ 的特征矩阵 $\lambda E-A$ 的 Smith 标准形.

解 $$\lambda E-A=\begin{pmatrix}\lambda-1 & -2 & -2\\ -2 & \lambda-1 & -2\\ -2 & -2 & \lambda-1\end{pmatrix}\sim\begin{pmatrix}\lambda-1 & -2 & -2\\ -1-\lambda & \lambda+1 & 0\\ -1-\lambda & 0 & \lambda+1\end{pmatrix}$$

$$\sim\begin{pmatrix}\lambda-5 & -2 & -2\\ 0 & \lambda+1 & 0\\ 0 & 0 & \lambda+1\end{pmatrix}\sim\begin{pmatrix}\lambda-5 & -2 & -2\\ 0 & \lambda+1 & \lambda+1\\ 0 & 0 & \lambda+1\end{pmatrix}$$

$$\sim\begin{pmatrix}\lambda-5 & -2 & 0\\ 0 & \lambda+1 & 0\\ 0 & 0 & \lambda+1\end{pmatrix}\sim\begin{pmatrix}0 & -2 & 0\\ \frac{1}{2}(\lambda+1)(\lambda-5) & \lambda+1 & 0\\ 0 & 0 & \lambda+1\end{pmatrix}$$

$$\sim\begin{pmatrix}0 & 1 & 0\\ \frac{1}{2}(\lambda+1)(\lambda-5) & 0 & 0\\ 0 & 0 & \lambda+1\end{pmatrix}\sim\begin{pmatrix}1 & 0 & 0\\ 0 & \lambda+1 & 0\\ 0 & 0 & (\lambda+1)(\lambda-5)\end{pmatrix}$$

【例 3.1.3】 将 λ-矩阵 $A(\lambda)=\begin{pmatrix}\lambda & 2\lambda+1 & 1\\ 1 & \lambda+1 & \lambda^2+1\\ \lambda-1 & \lambda & -\lambda^2\end{pmatrix}$ 化为 Smith 标准形.

解

$$A(\lambda)=\begin{pmatrix}\lambda & 2\lambda+1 & 1\\ 1 & \lambda+1 & \lambda^2+1\\ \lambda-1 & \lambda & -\lambda^2\end{pmatrix}\sim\begin{pmatrix}\lambda & 2\lambda+1 & 1\\ 1 & \lambda+1 & \lambda^2+1\\ \lambda & 2\lambda+1 & 1\end{pmatrix}\sim\begin{pmatrix}0 & -\lambda^2+\lambda+1 & 1-\lambda^3-\lambda\\ 1 & \lambda+1 & \lambda^2+1\\ 0 & 0 & 0\end{pmatrix}$$

$$\sim\begin{pmatrix}1 & 0 & 0\\ 0 & -\lambda^2+\lambda+1 & 1-\lambda^3-\lambda\\ 0 & 0 & 0\end{pmatrix}\sim\begin{pmatrix}1 & 0 & 0\\ 0 & -\lambda^2+\lambda+1 & 1-\lambda^2-2\lambda\\ 0 & 0 & 0\end{pmatrix}$$

$$\sim\begin{pmatrix}1&0&0\\0&\lambda&1-\lambda^2-2\lambda\\0&0&0\end{pmatrix}\sim\begin{pmatrix}1&0&0\\0&\lambda&1\\0&0&0\end{pmatrix}\sim\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}$$

§3.2 不变因子与初等因子

定义 3.2.1 设λ-矩阵$\boldsymbol{A}(\lambda)$的秩为r，对于$k\in\mathbf{Z}^+$，$1\leqslant k\leqslant r$，$\boldsymbol{A}(\lambda)$中必存在着非零的k阶子式．$\boldsymbol{A}(\lambda)$中全部k阶子式的最大公因式$\boldsymbol{D}_k(\lambda)$叫做$\boldsymbol{A}(\lambda)$的$k$阶行列式因子．

对于秩为r的λ-矩阵$\boldsymbol{A}(\lambda)$其非零主子式共有r个，并且其非零行列式因子也有r个．

定理 3.2.1 等价的λ-矩阵具有相同的秩和相同的各阶行列式因子．

证 要证明这一定理，只要能证明λ-矩阵$\boldsymbol{A}(\lambda)$经过一次初等变换，它的秩与行列式因子不变即可．

设λ-矩阵$\boldsymbol{A}(\lambda)$经过一次初等变换变成$\boldsymbol{B}(\lambda)$，$\boldsymbol{f}(\lambda)$与$\boldsymbol{g}(\lambda)$分别是$\boldsymbol{A}(\lambda)$与$\boldsymbol{B}(\lambda)$的k阶行列式因子，现在证明$\boldsymbol{f}(\lambda)=\boldsymbol{g}(\lambda)$．

情形一：$\boldsymbol{A}(\lambda)$经过交换两行或两列的初等变换变成$\boldsymbol{B}(\lambda)$，此时，$\boldsymbol{B}(\lambda)$的每个k阶子式或等于$\boldsymbol{A}(\lambda)$的某个k阶子式，或与$\boldsymbol{A}(\lambda)$的某一个k阶子式反号，因此$\boldsymbol{f}(\lambda)$是$\boldsymbol{B}(\lambda)$的k阶子式的公因式，所以$\boldsymbol{f}(\lambda)$可以整除$\boldsymbol{g}(\lambda)$．

情形二：$\boldsymbol{A}(\lambda)$经过以一个不为零的常数c乘某行或某列的初等变换变成$\boldsymbol{B}(\lambda)$，此时，$\boldsymbol{B}(\lambda)$的每个k阶子式或等于$\boldsymbol{A}(\lambda)$的某个k阶子式，或等于$\boldsymbol{A}(\lambda)$的某个k阶子式的c倍，因此$\boldsymbol{f}(\lambda)$是$\boldsymbol{B}(\lambda)$的k阶子式的公因式，所以$\boldsymbol{f}(\lambda)$可以整除$\boldsymbol{g}(\lambda)$．

情形三：$\boldsymbol{A}(\lambda)$经过以其第j行(或第j列)乘以多项式$\varphi(\lambda)$后加到第i行（或第i列）上的初等变换变成$\boldsymbol{B}(\lambda)$，此时，$\boldsymbol{B}(\lambda)$中那些包含i行与j行(或i列与j列)的k阶子式和那些不包含i行(或i列)的k阶子式等于$\boldsymbol{A}(\lambda)$中对应的k阶子式；$\boldsymbol{B}(\lambda)$中那些包含i行（或i列）但不包含j行（或j列）的k阶子式，按i行（或i列）分成两部分，等于$\boldsymbol{A}(\lambda)$的一个k阶子式与另一个的k阶子式的$\pm\varphi(\lambda)$倍的和，即$\boldsymbol{A}(\lambda)$的两个k阶子式的组合，因此$\boldsymbol{f}(\lambda)$是$\boldsymbol{B}(\lambda)$的k阶子式的公因式，所以$\boldsymbol{f}(\lambda)$可以整除$\boldsymbol{g}(\lambda)$．

由初等变换的可逆性，$\boldsymbol{B}(\lambda)$也可以经过一次初等变换变成$\boldsymbol{A}(\lambda)$，由相同的讨论可得$\boldsymbol{g}(\lambda)$可以整除$\boldsymbol{f}(\lambda)$．所以$\boldsymbol{f}(\lambda)=\boldsymbol{g}(\lambda)$．

如果$\boldsymbol{A}(\lambda)$的全部k阶子式都为零，有$\boldsymbol{f}(\lambda)=0$，所以$\boldsymbol{g}(\lambda)=0$，则$\boldsymbol{B}(\lambda)$的全部$k$阶子式等于零，反之亦然．所以$\boldsymbol{A}(\lambda)$和$\boldsymbol{B}(\lambda)$有相同的秩和相同的各阶行列式因子．

设λ-矩阵$\boldsymbol{A}(\lambda)$的 Smith 标准形为$\begin{pmatrix}\boldsymbol{d}_1(\lambda)&&&&&\\&\ddots&&&&\\&&\boldsymbol{d}_r(\lambda)&&&\\&&&0&&\\&&&&\ddots&\\&&&&&0\end{pmatrix}$，在这种形式的矩阵中，如果一个的$k$阶子式包含的行与列的标号不完全相同，那么这个$k$阶子式必为零，因此，由$i_1$，$i_2$，…，$i_k$行与$i_1$，$i_2$，…，$i_k$列（$1\leqslant i_1<i_2<\cdots<i_k\leqslant r$）组成的$k$阶子式$\boldsymbol{d}_{i_1}(\lambda)$，$\boldsymbol{d}_{i_2}(\lambda)$，…，$\boldsymbol{d}_{i_k}(\lambda)$就是矩阵$\boldsymbol{A}(\lambda)$的$k$阶行列式因子；而$\boldsymbol{d}_1(\lambda)\boldsymbol{d}_2(\lambda)\cdots\boldsymbol{d}_k(\lambda)$即为这

种 k 阶子式的最大公因式．因此 $A(\lambda)$ 的 k 阶行列式因子 $D_k = d_1(\lambda)d_2(\lambda)\cdots d_k(\lambda)$．

定理 3.2.2 λ-矩阵 $A(\lambda)$ 的 Smith 标准形是唯一的．

证 设 $A(\lambda)$ 的 Smith 标准形为 $\begin{pmatrix} d_1(\lambda) & & & & & \\ & \ddots & & & & \\ & & d_r(\lambda) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$，则 $A(\lambda)$ 与其等价，并且它们有相同的秩和相同的各阶行列式因子，所以 $A(\lambda)$ 的秩就是标准形主对角线上非零元素的个数 r；$A(\lambda)$ 的 k 阶行列式因子就是 $D_k(\lambda) = d_1(\lambda)d_2(\lambda)\cdots d_k(\lambda)\ (k = 1,2,\cdots,r)$ 所以 $d_1(\lambda) = D_1(\lambda)$，$d_2(\lambda) = \dfrac{D_2(\lambda)}{D_1(\lambda)},\cdots,d_r(\lambda) = \dfrac{D_r(\lambda)}{D_{r-1}(\lambda)}$；即 $d_i(\lambda)(i = 1,2,\cdots,r)$ 由 $A(\lambda)$ 的行列式因子唯一确定，所以，$A(\lambda)$ 的 Smith 标准形是唯一的．

定义 3.2.2 多项式矩阵 $A(\lambda)$ 的 Smith 标准形中的非零元素 $d_i(\lambda)(i = 1,2,\cdots,s)$ 叫做 $A(\lambda)$ 的不变因子（或不变因式）．

定理 3.2.3 两个 λ-矩阵等价的充分与必要条件为它们具有相同的行列式因子，或具有相同的不变因子．

证 因为不变因子与行列式因子是相互确定的，因此，必要性已经在定理 3.2.2 中被证明了.

充分性：若两个 λ-矩阵 $A(\lambda)$ 与 $B(\lambda)$ 有相同的不变因子，则 $A(\lambda)$、$B(\lambda)$ 和同一个标准形等价，所以 $A(\lambda)$ 与 $B(\lambda)$ 等价．

由于任意一个复数域上的多项式 $P(x)$，均可以分解为一次因式之积．对于 λ-矩阵 $A(\lambda)$ 而言，其不变因子亦可以被分解．假设其分解式为

$$\begin{cases} d_1(\lambda) = (\lambda-\lambda_1)^{e_{11}}(\lambda-\lambda_2)^{e_{12}}\cdots(\lambda-\lambda_s)^{e_{1s}} \\ d_2(\lambda) = (\lambda-\lambda_1)^{e_{21}}(\lambda-\lambda_2)^{e_{22}}\cdots(\lambda-\lambda_s)^{e_{2s}} \\ \cdots \\ d_r(\lambda) = (\lambda-\lambda_1)^{e_{r1}}(\lambda-\lambda_2)^{e_{r2}}\cdots(\lambda-\lambda_s)^{e_{rs}} \end{cases}$$

其中 λ_1，λ_2，…，λ_s 互不相同．

∵ $d_k(\lambda)$ 可以整除 $d_{k+1}(\lambda)$，所以 $e_{1j} \leqslant e_{2j} \leqslant \cdots \leqslant e_{rj}$（$j=1, 2, \cdots, s$），

∴ e_{r1}，e_{r2}，…，e_{rs} 全不为零．但是，e_{1j}，e_{2j}，…，$e_{r-1,j}$（$j=1, 2, \cdots, s$）中可能有零．如果某个 $e_{ij}=0$（$i=1, 2, \cdots, r-1$；$j=1, 2, \cdots, s$）则必定有 $e_{1j}=e_{2j}=\cdots=e_{i-1,j}=0$.

定义 3.2.3 λ-矩阵 $A(\lambda)$ 的分解式

$$\begin{cases} d_1(\lambda) = (\lambda-\lambda_1)^{e_{11}}(\lambda-\lambda_2)^{e_{12}}\cdots(\lambda-\lambda_s)^{e_{1s}} \\ d_2(\lambda) = (\lambda-\lambda_1)^{e_{21}}(\lambda-\lambda_2)^{e_{22}}\cdots(\lambda-\lambda_s)^{e_{2s}} \\ \cdots \\ d_r(\lambda) = (\lambda-\lambda_1)^{e_{r1}}(\lambda-\lambda_2)^{e_{r2}}\cdots(\lambda-\lambda_s)^{e_{rs}} \end{cases}$$

其中 λ_1，λ_2，…，λ_s 互不相同.

所有指数大于 0 的因子 $(\lambda-\lambda_j)^{e_{ij}}$（$i=1, 2, \cdots, r$；$j=1, 2, \cdots, s$；$e_{ij}>0$）叫做 $A(\lambda)$ 的初等因子．

定理 3.2.4 若 λ-矩阵 $A(\lambda)$ 的不变因子 $d_i(\lambda)$ 确定，则 $A(\lambda)$ 的初等因子 $(\lambda-\lambda_j)^{e_{ij}}$ 就被

唯一确定；反之，若所有的初等因子 $(\lambda-\lambda_j)^{e_{ij}}$ 与 $A(\lambda)$ 的秩确定，则不变因子 $d_i(\lambda)$ 也就被唯一确定.

证 将 $A(\lambda)$ 的所有初等因子按不同的因子分类，并按各因子的幂从小到大排列为

$$\begin{cases}(\lambda-\lambda_1)^{e_{11}},(\lambda-\lambda_1)^{e_{21}},\cdots,(\lambda-\lambda_1)^{e_{r1}}, & e_{11}\leqslant e_{21}\leqslant\cdots\leqslant e_{r1},\\ (\lambda-\lambda_2)^{e_{12}},(\lambda-\lambda_2)^{e_{22}},\cdots,(\lambda-\lambda_2)^{e_{r2}}, & e_{12}\leqslant e_{22}\leqslant\cdots\leqslant e_{r2}\\ \quad\vdots & \quad\vdots\\ (\lambda-\lambda_s)^{e_{1s}},(\lambda-\lambda_s)^{e_{2s}},\cdots,(\lambda-\lambda_s)^{e_{rs}}, & e_{1s}\leqslant e_{2s}\leqslant\cdots\leqslant e_{rs}\end{cases}$$

此表中共有 r 列，每列中的因子个数可能不同，空白处可以以数 1 补上．这样，在第 i 列上的各式之积 $(\lambda-\lambda_1)^{e_{i1}}(\lambda-\lambda_2)^{e_{i2}}\cdots(\lambda-\lambda_s)^{e_{is}}$ 就是 $A(\lambda)$ 的第 i 个不变因子 $d_i(\lambda)$，所以，当 $A(\lambda)$ 的秩确定后，求 $A(\lambda)$ 不变因子等于求它的初等因子.

定理 3.2.5 设 λ-矩阵 $A(\lambda)=\begin{pmatrix}B(\lambda) & 0\\ 0 & C(\lambda)\end{pmatrix}$ 为分块对角形矩阵，则 $B(\lambda)$ 与 $C(\lambda)$ 的初等因子的全体构成 $A(\lambda)$ 的全体初等因子.

证 首先将 $B(\lambda)$ 与 $C(\lambda)$ 化为与之等价的 Smith 标准形

$$B(\lambda)=\begin{bmatrix}b_1(\lambda) &&&&&\\ &\ddots&&&&\\ &&b_{r_1}(\lambda)&&&\\ &&&0&&\\ &&&&\ddots&\\ &&&&&0\end{bmatrix},\quad C(\lambda)=\begin{bmatrix}c_1(\lambda) &&&&&\\ &\ddots&&&&\\ &&c_{r_2}(\lambda)&&&\\ &&&0&&\\ &&&&\ddots&\\ &&&&&0\end{bmatrix}$$

显然 $\operatorname{rank}[A(\lambda)]=r=\operatorname{rank}[B(\lambda)]+\operatorname{rank}[C(\lambda)]=r_1+r_2$.

再将 $b_k(\lambda)$ 与 $c_l(\lambda)$ 分解为因式之积

$$b_k(\lambda)=(\lambda-\lambda_1)^{e_{k1}}(\lambda-\lambda_2)^{e_{k2}}\cdots(\lambda-\lambda_k)^{e_{kt}},\quad k=1,2,\cdots,r_1$$

$$c_l(\lambda)=(\lambda-\lambda_1)^{i_{l1}}(\lambda-\lambda_2)^{i_{l2}}\cdots(\lambda-\lambda_k)^{i_{lt}},\quad l=1,2,\cdots,r_2$$

以下证明 $B(\lambda)$ 与 $C(\lambda)$ 的初等因子就是 $A(\lambda)$ 的全部初等因子.

不失一般性，考虑 $B(\lambda)$ 与 $C(\lambda)$ 中只含因子 $(\lambda-\lambda_1)$ 的那些初等因子，将因子 $(\lambda-\lambda_1)$ 的指数 e_{11}，e_{21}，$\cdots$，$e_{r_1 1}$，i_{11}，i_{21}，$\cdots$，$i_{r_2 1}$ 按大小排列，记为 j_1，j_2，$\cdots$，j_r，满足 $0\leqslant j_1\leqslant j_2\leqslant\cdots\leqslant j_r$. 对 $A(\lambda)=\begin{pmatrix}B(\lambda) & 0\\ 0 & C(\lambda)\end{pmatrix}$ 作初等变换，得到

$$A(\lambda)\sim\begin{bmatrix}b_1(\lambda)&&&&&&&&\\ &\ddots&&&&&&&\\ &&b_{r_1}(\lambda)&&&&&&\\ &&&c_1(\lambda)&&&&&\\ &&&&\ddots&&&&\\ &&&&&c_{r_2}(\lambda)&&&\\ &&&&&&0&\cdots&0\\ &&&&&&\vdots&\ddots&\vdots\\ &&&&&&0&\cdots&0\end{bmatrix}$$

$$\sim\begin{pmatrix} (\lambda-\lambda_1)^{j_1}\boldsymbol{\varphi}_1(\lambda) & & & & & & \\ & (\lambda-\lambda_1)^{j_2}\boldsymbol{\varphi}_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & (\lambda-\lambda_1)^{j_r}\boldsymbol{\varphi}_r(\lambda) & & & \\ & & & & 0 & \cdots & 0 \\ & & & & \vdots & \ddots & \vdots \\ & & & & 0 & \cdots & 0 \end{pmatrix}$$

其中 $\boldsymbol{\varphi}_1(\lambda)$，$\boldsymbol{\varphi}_2(\lambda)$，…，$\boldsymbol{\varphi}_r(\lambda)$ 中均不含因子 $(\lambda-\lambda_1)$.

所以在 $\boldsymbol{A}(\lambda)$ 的 k 阶行列式因子 $\boldsymbol{D}_1(\lambda)$，$\boldsymbol{D}_2(\lambda)$，…，$\boldsymbol{D}_r(\lambda)$ 中，$(\lambda-\lambda_1)$ 的幂指数分别为 j_1，j_2，…，j_r；即 $\boldsymbol{A}(\lambda)$ 中含有因子 $(\lambda-\lambda_1)$ 的初等因子为 $(\lambda-\lambda_1)^{j_k}$ $(j_k\neq 0$；$k=1$，2，…，$r)$．对于因子 $(\lambda-\lambda_2)$，$(\lambda-\lambda_3)$，…，$(\lambda-\lambda_t)$ 可以类似地进行讨论．

最后证明，除去 $\boldsymbol{B}(\lambda)$ 与 $\boldsymbol{C}(\lambda)$ 的初等因子之外，$\boldsymbol{A}(\lambda)$ 没有其他的初等因子，因为 $\mathrm{rank}\boldsymbol{A}(\lambda)=r$，所以 $\boldsymbol{D}_r(\lambda)=\boldsymbol{b}_1(\lambda)\cdots\boldsymbol{b}_{r_1}(\lambda)\boldsymbol{c}_1(\lambda)\cdots\boldsymbol{c}_{r_2}(\lambda)$ 为 $\boldsymbol{A}(\lambda)$ 的全部初等因子之积，如果 $(\lambda-a)^k$ 是 $\boldsymbol{A}(\lambda)$ 的一个初等因子，则它必然包含在某个 $\boldsymbol{b}_i(\lambda)$ 或 $\boldsymbol{c}_j(\lambda)$ 之中．因此，$\boldsymbol{B}(\lambda)$ 与 $\boldsymbol{C}(\lambda)$ 的初等因子的全体构成 $\boldsymbol{A}(\lambda)$ 的全体初等因子．

一般的，如果 $\boldsymbol{A}(\lambda)=\begin{pmatrix}\boldsymbol{B}_1(\lambda) & & & & \\ & \boldsymbol{B}_2(\lambda) & & & \\ & & \ddots & & \\ & & & \boldsymbol{B}_t(\lambda) & \\ & & & & 0\end{pmatrix}$，则 $\boldsymbol{B}_i(\lambda)$ $(i=1, 2, \cdots, t)$ 的初等因子的全体构成 $\boldsymbol{A}(\lambda)$ 的全体初等因子．

【例 3.2.1】　设 $\boldsymbol{A}(\lambda)$ 为 5 阶方阵，其秩为 4，全体初等因子是

$$\lambda,\lambda^2,\lambda^2,\lambda-1,\lambda-1,\lambda+1,(\lambda+1)^2$$

试求 $\boldsymbol{A}(\lambda)$ 的不变因子和 Smith 标准形．

解　$\boldsymbol{d}_1(\lambda)=1$，$\boldsymbol{d}_2(\lambda)=\lambda$，$\boldsymbol{d}_3(\lambda)=\lambda^2(\lambda-1)(\lambda+1)$，$\boldsymbol{d}_4(\lambda)=\lambda^2(\lambda-1)(\lambda+1)^2$

$\boldsymbol{A}(\lambda)$ 的 Smith 标准形为

$$\begin{pmatrix}1 & 0 & 0 & 0 & 0\\ 0 & \lambda & 0 & 0 & 0\\ 0 & 0 & \lambda^2(\lambda-1)(\lambda+1) & 0 & 0\\ 0 & 0 & 0 & \lambda^2(\lambda-1)(\lambda+1)^2 & 0\\ 0 & 0 & 0 & 0 & 0\end{pmatrix}$$

【例 3.2.2】　将 λ-矩阵 $\boldsymbol{A}(\lambda)=\begin{pmatrix}0 & 0 & 0 & \lambda^2\\ 0 & 0 & \lambda^2-\lambda & 0\\ 0 & (\lambda-1)^2 & 0 & 0\\ \lambda^2-\lambda & 0 & 0 & 0\end{pmatrix}$ 化为 Smith 标准形．

解法一　因为 $\boldsymbol{A}(\lambda)$ 的各阶行列式因子为

$$\boldsymbol{D}_1(\lambda)=1,\boldsymbol{D}_2(\lambda)=\lambda(\lambda-1),\boldsymbol{D}_3(\lambda)=\lambda^2(\lambda-1)^2,\boldsymbol{D}_4(\lambda)=\lambda^4(\lambda-1)^4$$

所以 $\boldsymbol{A}(\lambda)$ 的不变因子为

$$d_1(\lambda)=1,\quad d_2(\lambda)=\frac{D_2(\lambda)}{D_1(\lambda)}=\lambda(\lambda-1),d_3(\lambda)=\frac{D_3(\lambda)}{D_2(\lambda)}=\lambda(\lambda-1)$$

$$d_4(\lambda)=\frac{D_4(\lambda)}{D_3(\lambda)}=\lambda^2(\lambda-1)^2$$

因此，$A(\lambda)$ 的 Smith 标准形为

$$\begin{bmatrix}1&0&0&0\\0&\lambda(\lambda-1)&0&0\\0&0&\lambda(\lambda-1)&0\\0&0&0&\lambda^2(\lambda-1)^2\end{bmatrix}$$

解法二 由次对角线上的元素可得 $A(\lambda)$ 的初等因子为

$$\lambda^2,\lambda,\lambda-1,(\lambda-1)^2,\lambda,\lambda-1$$

所以 $A(\lambda)$ 的不变因子为

$$d_4(\lambda)=\lambda^2(\lambda-1)^2,d_3(\lambda)=\lambda(\lambda-1),d_2(\lambda)=\lambda(\lambda-1),d_1(\lambda)=1$$

因此，$A(\lambda)$ 的 Smith 标准形为

$$\begin{bmatrix}1&0&0&0\\0&\lambda(\lambda-1)&0&0\\0&0&\lambda(\lambda-1)&0\\0&0&0&\lambda^2(\lambda-1)^2\end{bmatrix}$$

【例 3.2.3】 求 λ-矩阵 $A(\lambda)=\begin{bmatrix}\lambda(\lambda+1)&&\\&\lambda&\\&&(\lambda+1)^2\end{bmatrix}$ 的初等因子和不变因子．

解法一 利用行列式因子．

∵ λ 与 $(\lambda+1)^2$ 的最大公因式为 1，所以 $D_1(\lambda)=1$；

又∵

$$D_2(\lambda)=\lambda(\lambda+1),D_3(\lambda)=\lambda^2(\lambda+1)^3$$

∴ $A(\lambda)$ 的不变因子为 1，$\lambda(\lambda+1)$，$\lambda(\lambda+1)^2$，其初等因子为：λ，$\lambda+1$，λ，$(\lambda+1)^2$．

解法二 因为 $A(\lambda)$ 是一个对角形矩阵，对角线上各元素的初等因子就是 $A(\lambda)$ 的初等因子，即 λ，$\lambda+1$，λ，$(\lambda+1)^2$；而不变因子为 $d_3(\lambda)=\lambda(\lambda+1)^2,d_2(\lambda)=\lambda(\lambda+1),d_1(\lambda)=1$.

【例 3.2.4】 将 λ-矩阵 $A(\lambda)=\begin{bmatrix}3\lambda+5&(\lambda+2)^2&4\lambda+5&(\lambda-1)^2\\\lambda+7&(\lambda+2)^2&\lambda+7&0\\\lambda-1&0&2\lambda-1&(\lambda-1)^2\\0&0&\lambda&0\end{bmatrix}$ 化为 Smith 标准形．

解 对 λ-矩阵 $A(\lambda)$ 进行初等变换，可得

$$\begin{bmatrix}3\lambda+5&(\lambda+2)^2&4\lambda+5&(\lambda-1)^2\\\lambda+7&(\lambda+2)^2&\lambda+7&0\\\lambda-1&0&2\lambda-1&(\lambda-1)^2\\0&0&\lambda&0\end{bmatrix}\sim\begin{bmatrix}2\lambda+6&(\lambda+2)^2&2\lambda+6&0\\\lambda+7&(\lambda+2)^2&\lambda+7&0\\\lambda-1&0&2\lambda-1&(\lambda-1)^2\\0&0&\lambda&0\end{bmatrix}$$

$$\sim \begin{pmatrix} \lambda-1 & 0 & 0 & 0 \\ \lambda+7 & (\lambda+2)^2 & 0 & 0 \\ \lambda-1 & 0 & \lambda & (\lambda-1)^2 \\ 0 & 0 & \lambda & 0 \end{pmatrix} \sim \begin{pmatrix} \lambda-1 & 0 & 0 & 0 \\ 6 & (\lambda+2)^2 & 0 & 0 \\ 0 & 0 & \lambda & (\lambda-1)^2 \\ 0 & 0 & \lambda & 0 \end{pmatrix} \sim \begin{pmatrix} \boldsymbol{A}_1(\lambda) & 0 \\ 0 & \boldsymbol{A}_2(\lambda) \end{pmatrix}$$

$\boldsymbol{A}_1(\lambda)=\begin{pmatrix} \lambda-1 & 0 \\ 6 & (\lambda+2)^2 \end{pmatrix}$ 的初等因子为 $\lambda-1$，$(\lambda+2)^2$，$\boldsymbol{A}_2(\lambda)=\begin{pmatrix} \lambda & 0 \\ \lambda & (\lambda-1)^2 \end{pmatrix}$ 的初等因子为 λ，$(\lambda-1)^2$，因此 $\boldsymbol{A}(\lambda)$ 的初等因子为 $\lambda, (\lambda-1)^2, \lambda-1, (\lambda+2)^2$．易见，$\boldsymbol{A}(\lambda)$ 的秩是 4.

因此，$\boldsymbol{A}(\lambda)$ 的 Smith 标准形为 $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \lambda-1 & 0 \\ 0 & 0 & 0 & \lambda(\lambda-1)^2(\lambda+2)^2 \end{pmatrix}$

【例 3.2.5】 试求矩阵 $\boldsymbol{A}=\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{pmatrix}$ 的特征矩阵 $\lambda\boldsymbol{E}-\boldsymbol{A}$ 的不变因子，并将其化为 Smith 标准形．

解 因为 $\boldsymbol{A}$ 的特征矩阵是 $\lambda\boldsymbol{E}-\boldsymbol{A}=\begin{pmatrix} \lambda & -1 & 0 & \cdots & 0 \\ 0 & \lambda & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \\ a_n & a_{n-1} & a_{n-2} & \cdots & \lambda+a_1 \end{pmatrix}$，其右上角的 $n-1$ 阶子式为 $(-1)^{n-1}$，因此其 $n-1$ 阶行列式因子为 $\boldsymbol{D}_{n-1}(\lambda)=1$，从而 $\boldsymbol{d}_{n-1}(\lambda)=1$．而

$$\boldsymbol{d}_n(\lambda)=\frac{\boldsymbol{D}_n(\lambda)}{\boldsymbol{d}_{n-1}(\lambda)}=\lambda^n+a_1\lambda^{n-1}+\cdots+a_{n-1}\lambda+a_n$$

因此，$\lambda\boldsymbol{E}-\boldsymbol{A}$ 的不变因子为 $\lambda^n+a_1\lambda^{n-1}+\cdots+a_{n-1}\lambda+a_n$，Smith 标准形为

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \lambda^n+a_1\lambda^{n-1}+\cdots+a_{n-1}\lambda+a_n \end{pmatrix}$$

【例 3.2.6】 试求矩阵 $\boldsymbol{A}(\lambda)=\begin{pmatrix} \lambda-a & c_1 & & & \\ & \lambda-a & c_2 & & \\ & & \ddots & \ddots & \\ & & & \lambda-a & c_{n-1} \\ & & & & \lambda-a \end{pmatrix}$ 的不变因子与初等因子，其中 c_1，c_2，…，c_{n-1} 为非零常数．

解 因为 $\boldsymbol{A}(\lambda)$ 为 n 阶方阵，所以 $\boldsymbol{D}_n(\lambda)=(\lambda-a)^n$．去掉第 n 行与第 1 列后，余下的 $n-1$ 阶子式为 $c_1c_2\cdots c_{n-1}$，所以 $\boldsymbol{D}_{n-1}(\lambda)=1$．因此 $\boldsymbol{D}_1(\lambda)=\boldsymbol{D}_2(\lambda)=\cdots=\boldsymbol{D}_{n-1}(\lambda)=1$，从而 $\boldsymbol{A}(\lambda)$ 的不变因子为

$$d_1(\lambda)=d_2(\lambda)=\cdots=d_{n-1}(\lambda)=1, d_n(\lambda)=(\lambda-a)^n$$

$A(\lambda)$ 的初等因子为 $(\lambda-a)^n$.

下面将利用 λ-矩阵的等价来刻画两个矩阵的相似.

定理 3.2.6 设 A, B 是 n 阶方阵，则 A 与 B 相似的充分且必要条件为它们的特征矩阵 $\lambda E-A\sim\lambda E-B$.

证 必要性 设 A 与 B 相似，则有可逆数值矩阵 P，使得 $A=P^{-1}BP$. 于是

$$\lambda E-A=\lambda E-P^{-1}BP=P^{-1}(\lambda E-A)P$$

从而 $\lambda E-A\sim\lambda E-B$.

充分性的证明略.

注意到 $\lambda E-A$ 的不变因子与 $\lambda E-A$ 的初等因子的关系，则得下面推论.

推论 3.2.1 矩阵 A 与 B 相似的充分且必要条件是它们有相同的不变因子，有相同的 Smith 标准形.

【例 3.2.7】 在下列矩阵对中，哪组矩阵是相似的，哪组是不相似的：

$$(1) A_1=\begin{pmatrix}2&1&0&0\\0&2&1&0\\0&0&2&0\\0&0&0&3\end{pmatrix},\quad A_2=\begin{pmatrix}2&0&0&0\\0&2&1&0\\0&0&2&0\\0&0&0&3\end{pmatrix};$$

$$(2) B_1=\begin{pmatrix}2&0&0&0\\0&2&1&0\\0&0&2&0\\0&0&0&3\end{pmatrix},\quad B_2=\begin{pmatrix}2&1&0&0\\0&2&0&0\\0&0&2&0\\0&0&0&3\end{pmatrix};$$

$$(3) C_1=\begin{pmatrix}2&1&0&0\\0&2&0&0\\0&0&2&0\\0&0&0&3\end{pmatrix},\quad C_2=\begin{pmatrix}2&1&0&0\\0&2&0&0\\0&0&2&1\\0&0&0&3\end{pmatrix}.$$

解 (1)

$$\lambda E-A_1=\begin{pmatrix}\lambda-2&-1&0&0\\0&\lambda-2&-1&0\\0&0&\lambda-2&0\\0&0&0&\lambda-3\end{pmatrix}\sim\begin{pmatrix}1&&&\\&1&&\\&&1&\\&&&(\lambda-2)^3(\lambda-3)\end{pmatrix}$$

$$\lambda E-A_2=\begin{pmatrix}\lambda-2&0&0&0\\0&\lambda-2&-1&0\\0&0&\lambda-2&0\\0&0&0&\lambda-3\end{pmatrix}\sim\begin{pmatrix}1&&&\\&1&&\\&&\lambda-2&\\&&&(\lambda-2)^2(\lambda-3)\end{pmatrix}$$

因为 $\lambda E-A_1$ 与 $\lambda E-A_2$ 有不同的 Smith 标准型，因此 A_1 与 A_2 不相似.

(2)

$$\lambda E-B_1=\begin{pmatrix}\lambda-2&0&0&0\\0&\lambda-2&-1&0\\0&0&\lambda-2&0\\0&0&0&\lambda-3\end{pmatrix}\sim\begin{pmatrix}1&&&\\&1&&\\&&\lambda-2&\\&&&(\lambda-2)^2(\lambda-3)\end{pmatrix}$$

$$\lambda E-B_2=\begin{pmatrix}\lambda-2 & -1 & 0 & 0\\ 0 & \lambda-2 & 0 & 0\\ 0 & 0 & \lambda-2 & 0\\ 0 & 0 & 0 & \lambda-3\end{pmatrix}\sim\begin{pmatrix}1 & & & \\ & 1 & & \\ & & \lambda-2 & \\ & & & (\lambda-2)^2(\lambda-3)\end{pmatrix}$$

因为 $\lambda E-B_1$ 与 $\lambda E-B_2$ 有相同的 Smith 标准形，因此 B_1 与 B_2 相似．

(3)

$$\lambda E-C_1=\begin{pmatrix}\lambda-2 & -1 & 0 & 0\\ 0 & \lambda-2 & 0 & 0\\ 0 & 0 & \lambda-2 & 0\\ 0 & 0 & 0 & \lambda-3\end{pmatrix}\sim\begin{pmatrix}1 & & & \\ & 1 & & \\ & & \lambda-2 & \\ & & & (\lambda-2)^2(\lambda-3)\end{pmatrix}$$

$$\lambda E-C_2=\begin{pmatrix}\lambda-2 & -1 & 0 & 0\\ 0 & \lambda-2 & 0 & 0\\ 0 & 0 & \lambda-2 & -1\\ 0 & 0 & 0 & \lambda-3\end{pmatrix}\sim\begin{pmatrix}1 & & & \\ & 1 & & \\ & & \lambda-2 & \\ & & & (\lambda-2)^2(\lambda-3)\end{pmatrix}$$

因为 $\lambda E-C_1$ 与 $\lambda E-C_2$ 有相同的 Smith 标准形，因此 C_1 与 C_2 相似．

§3.3 Jordan 标准形

n 阶矩阵可以分为许多相似类，我们希望在与某个 n 阶矩阵相似的全体矩阵中，找到一个比较简单的矩阵，作为这一类矩阵的代表，从而简化这一类矩阵的讨论．当然对角形矩阵最为简单，但是并不是任意的 n 阶矩阵都能与对角矩阵相似的，例如非单纯矩阵就不与任意对角矩阵相似，但是却相似于一个我们称之为 Jordan 标准形的矩阵，这是一个相对简单的矩阵．

定义 3.3.1 形如 $J_i=\begin{pmatrix}\lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \lambda_i & \ddots & \\ & & & \ddots & 1\\ & & & & \lambda_i\end{pmatrix}_{m_i\times m_i}$ 的方阵叫做 m_i 阶 Jordan 块．特别的，一阶方阵叫做一阶 Jordan 块．

定义 3.3.2 由若干个 Jordan 块组成的分块对角矩阵 $\begin{pmatrix}J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_t\end{pmatrix}$ 其中 J_i $(i=1,2,\cdots,t)$ 为 m_i 阶 Jordan 块，当 $\sum\limits_{i=1}^{t}m_i=n$ 时，这个矩阵叫做 n 阶 Jordan 标准形，记为 J.

由例 3.2.6 可知 m_i 阶 Jordan 块 J_i 的特征矩阵 $\lambda E-J_i$ 的初等因子是 $(\lambda-\lambda_i)^{m_i}$，所以，

Jordan 标准形 $\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_t \end{pmatrix}$ 的特征矩阵$\lambda E-J$ 的初等因子是由各个 Jordan 块的初等因子合在一起构成的，即为 $(\lambda-\lambda_1)^{m_1}$，$(\lambda-\lambda_2)^{m_2}$，…，$(\lambda-\lambda_t)^{m_t}$，并且 $\sum_{i=1}^{t} m_i = n$.

因此，如果不计 J 中的 Jordan 块的排列顺序，那么 Jordan 标准形就由它的全部初等因子所决定.

定理 3.3.1 任意一个 n 阶复矩阵 A 都与一个 Jordan 标准形 J 相似，若不计 J 中的 Jordan 块的排列顺序，则 J 由 A 唯一确定.

证 设 A 的特征矩阵 $\lambda E-A$ 的初等因子为 $(\lambda-\lambda_1)^{m_1}$，$(\lambda-\lambda_2)^{m_2}$，…，$(\lambda-\lambda_t)^{m_t}$，其中 $\sum_{i=1}^{t} m_i = n$. 而每一个 $(\lambda-\lambda_i)^{m_i}$ 对应着一个 Jordan 块 J_i，其阶数为 m_i，主对角线上的元素为 λ_i，所有的 J_i 构成 J. 所以，$\lambda E-J$ 的全部初等因子就是 $(\lambda-\lambda_1)^{m_1}$，$(\lambda-\lambda_2)^{m_2}$，…，$(\lambda-\lambda_t)^{m_t}$.

由于 $\lambda E-J$ 与 $\lambda E-A$ 有相同的初等因子，因此 $\lambda E-J$ 与 $\lambda E-A$ 相似，由推论 3.2.6，J 与 A 相似.

若存在 J，$\tilde{J}$，均与 A 相似，则 $\lambda E-J$ 与 $\lambda E-\tilde{J}$ 有相同的初等因子，如果不计 J 与 $\tilde{J}$ 中的 Jordan 块的排列顺序，则有 $J=\tilde{J}$.

推论 3.3.1 任意一个 n 阶复矩阵 A 可以对角化的充分且必要条件是 A 的特征矩阵 $\lambda E-A$ 的初等因子全是一次的.

定义 3.3.3 设 A 是 n 阶方阵，称 A 的特征矩阵 $\lambda E-A$ 的初等因子为 A 的初等因子.

应该指出，对 n 阶方阵 A，它的特征矩阵的秩一定是 n. 因此 A 的不变因子总是有 n 个，并且它们的乘积就等于这个矩阵的特征多项式.

【例 3.3.1】 设 A 的初等因子是 λ^2，λ^2，$\lambda-1$，$(\lambda+2)^2$，写出 A 的 Jordan 标准形 J.

解

$$J=\begin{pmatrix} 0 & 1 & & & & & \\ & 0 & & & & & \\ & & 0 & 1 & & & \\ & & & 0 & & & \\ & & & & 1 & & \\ & & & & & -2 & 1 \\ & & & & & & -2 \end{pmatrix}$$

【例 3.3.2】 在复数域上求矩阵 $A=\begin{pmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & -5 \end{pmatrix}$ 的 Jordan 标准形 J.

解 $\because \lambda E-A=\begin{pmatrix} \lambda-3 & 0 & -8 \\ -3 & \lambda+1 & -6 \\ 2 & 0 & \lambda+5 \end{pmatrix} \sim \begin{pmatrix} 1 & & \\ & \lambda+1 & \\ & & (\lambda+1)^2 \end{pmatrix}$，

∴初等因子为$\lambda+1$，$(\lambda+1)^2$，所以，A 的 Jordan 标准形 $J=\begin{pmatrix}-1&0&0\\0&-1&1\\0&0&-1\end{pmatrix}$.

【例 3.3.3】 在复数域上求矩阵 $A=\begin{pmatrix}8&-3&6\\3&-2&0\\-4&2&-2\end{pmatrix}$ 的 Jordan 标准形 J.

解 因为 $\lambda E-A=\begin{pmatrix}\lambda-8&3&-6\\-3&\lambda+2&0\\4&-2&\lambda+2\end{pmatrix}$ 中有两个 2 阶子式

$$\begin{vmatrix}3&-6\\\lambda+2&0\end{vmatrix}=6(\lambda+2),\begin{vmatrix}-3&\lambda+2\\4&-2\end{vmatrix}=-4\lambda-2$$

的最大公因子是 1，因此 $D_2(\lambda)=1$，而 $|\lambda E-A|=(\lambda-1)^2(\lambda-2)$，因此

$$\lambda E-A\sim\begin{pmatrix}1&&\\&1&\\&&(\lambda-1)^2(\lambda-2)\end{pmatrix}$$

从而 A 的 Jordan 标准形 $J=\begin{pmatrix}1&1&0\\0&1&0\\0&0&2\end{pmatrix}$.

【例 3.3.4】 在复数域上求矩阵 $A=\begin{pmatrix}3&1&0&0\\-4&-1&0&0\\0&0&2&1\\0&0&-1&0\end{pmatrix}$ 的 Jordan 标准形 J.

解 将矩阵 A 分块为矩阵 $A=\begin{pmatrix}A_1&0\\0&A_2\end{pmatrix}$，其中 $A_1=\begin{pmatrix}3&1\\-4&-1\end{pmatrix}$，$A_2=\begin{pmatrix}2&1\\-1&0\end{pmatrix}$，然后分别求出 A_1，A_2 的 Jordan 标准形，再组合成 A 的 Jordan 标准形（请读者自己完成）.

【例 3.3.5】 求矩阵 $A=\begin{pmatrix}-1&2&2&0\\3&-1&1&0\\2&2&-1&0\\1&-4&3&3\end{pmatrix}$ 的 Jordan 标准形 J_A.

解
$$|A-\lambda E|=\begin{vmatrix}-1-\lambda&2&2&0\\3&-1-\lambda&1&0\\2&2&-1-\lambda&0\\1&-4&3&3-\lambda\end{vmatrix}$$

$$=(3-\lambda)\begin{vmatrix}-1-\lambda&2&2\\3&-1-\lambda&1\\2&2&-1-\lambda\end{vmatrix}$$

$$=(3-\lambda)\begin{vmatrix}3-\lambda&2&2\\3-\lambda&-1-\lambda&1\\3-\lambda&2&-1-\lambda\end{vmatrix}$$

$$=(3-\lambda)^2\begin{vmatrix}1 & 2 & 2\\ 1 & -1-\lambda & 1\\ 1 & 2 & -1-\lambda\end{vmatrix}$$

$$=(3-\lambda)^2\begin{vmatrix}1 & 2 & 2\\ 0 & -3-\lambda & -1\\ 0 & 0 & -3-\lambda\end{vmatrix}=(3-\lambda)^2(3+\lambda)^2.$$

因此，$\boldsymbol{A}$ 的初等因式是如下之一：

(1) $\lambda-3$，$\lambda-3$，$\lambda+3$，$\lambda+3$；

(2) $\lambda-3$，$\lambda-3$，$(\lambda+3)^2$；

(3) $\lambda+3$，$\lambda+3$，$(\lambda-3)^2$；

(4) $(\lambda-3)^2$，$(\lambda+3)^2$.

从而 $\boldsymbol{A}$ 的 Jordan 标准形是如下之一

$$\boldsymbol{J}_1=\begin{pmatrix}3 & 0 & 0 & 0\\ 0 & 3 & 0 & 0\\ 0 & 0 & -3 & 0\\ 0 & 0 & 0 & -3\end{pmatrix},\quad \boldsymbol{J}_2=\begin{pmatrix}3 & 0 & 0 & 0\\ 0 & 3 & 0 & 0\\ 0 & 0 & -3 & 1\\ 0 & 0 & 0 & -3\end{pmatrix},$$

$$\boldsymbol{J}_3=\begin{pmatrix}3 & 1 & 0 & 0\\ 0 & 3 & 0 & 0\\ 0 & 0 & -3 & 0\\ 0 & 0 & 0 & -3\end{pmatrix},\quad \boldsymbol{J}_4=\begin{pmatrix}3 & 1 & 0 & 0\\ 0 & 3 & 0 & 0\\ 0 & 0 & -3 & 1\\ 0 & 0 & 0 & -3\end{pmatrix}$$

由于 $\boldsymbol{A}$ 与 $\boldsymbol{J}_A$ 相似，从而 $\boldsymbol{A}-k\boldsymbol{E}$ 与 $k\boldsymbol{E}-\boldsymbol{J}_A$ 相似，因此 $\boldsymbol{A}-k\boldsymbol{E}$ 与 $k\boldsymbol{E}-\boldsymbol{J}_A$ 有相同的秩.

注意到

$$\boldsymbol{A}-3\boldsymbol{E}=\begin{pmatrix}-4 & 2 & 2 & 0\\ 3 & -4 & 1 & 0\\ 2 & 2 & -4 & 0\\ 1 & -4 & 3 & 0\end{pmatrix}\sim\begin{pmatrix}1 & -4 & 3 & 0\\ 3 & -4 & 1 & 0\\ 2 & 2 & -4 & 0\\ -4 & 2 & 2 & 0\end{pmatrix}\sim\begin{pmatrix}1 & -4 & 3 & 0\\ 0 & 8 & -8 & 0\\ 0 & 10 & -10 & 0\\ 0 & -14 & 14 & 0\end{pmatrix}$$

$$\boldsymbol{A}+3\boldsymbol{E}=\begin{pmatrix}2 & 2 & 2 & 0\\ 3 & 2 & 1 & 0\\ 2 & 2 & 2 & 0\\ 1 & -4 & 3 & 6\end{pmatrix}\sim\begin{pmatrix}1 & 1 & 1 & 0\\ 3 & 2 & 1 & 0\\ 0 & 0 & 0 & 0\\ 1 & -4 & 3 & 6\end{pmatrix}$$

因此 $\boldsymbol{A}-3\boldsymbol{E}$ 的秩是 2，$\boldsymbol{A}+3\boldsymbol{E}$ 的秩是 3. 在 $\boldsymbol{J}_1$，$\boldsymbol{J}_2$，$\boldsymbol{J}_3$，$\boldsymbol{J}_4$，满足 $\boldsymbol{J}_i-3\boldsymbol{E}$ 的秩是 2，$\boldsymbol{J}_i+3\boldsymbol{E}$ 的秩是 3 的只有 $\boldsymbol{J}_2$，因此 $\boldsymbol{J}_A=\boldsymbol{J}_2$.

对 $\lambda\boldsymbol{E}-\boldsymbol{J}_i=\begin{pmatrix}\lambda-\lambda_i & -1 & & & \\ & \lambda-\lambda_i & -1 & & \\ & & \lambda-\lambda_i & \ddots & \\ & & & \ddots & -1\\ & & & & \lambda-\lambda_i\end{pmatrix}_{m_i\times m_i}$ 其秩有如下结论

$$\operatorname{rank}((\lambda\boldsymbol{E}-\boldsymbol{J}_i)^k)=\begin{cases}m_i-k, & \lambda=\lambda_i,k\leqslant m_i\\ 0, & \lambda=\lambda_i,k>m_i\\ m_i, & \lambda\neq\lambda_i\end{cases}$$

因此，若以 a 为特征值的 i 阶 Jordan 块的个数是 k_i，$i=1, 2, \cdots, s$，则

$$r_1(a) = \text{rank}(a\boldsymbol{E}-\boldsymbol{J}) = n-k_1-k_2-\cdots-k_s$$
$$r_2(a) = \text{rank}((a\boldsymbol{E}-\boldsymbol{J})^2) = n-k_1-2k_2-\cdots-2k_s$$
$$r_3(a) = \text{rank}((a\boldsymbol{E}-\boldsymbol{J})^3) = n-k_1-2k_2-3k_3-\cdots-3k_s$$
$$\vdots$$
$$r_s(a) = \text{rank}((a\boldsymbol{E}-\boldsymbol{J})^s) = n-k_1-2k_2-\cdots-sk_s$$
$$r_{s+m}(a) = \text{rank}((a\boldsymbol{E}-\boldsymbol{J})^{s+m}) = n-k_1-2k_2-\cdots-sk_s, m=1,2,\cdots$$

因为 $\boldsymbol{P}^{-1}(\lambda\boldsymbol{E}-\boldsymbol{A})^k\boldsymbol{P}=(\lambda\boldsymbol{E}-\boldsymbol{J})^k$，所以 $\text{rank}((\lambda\boldsymbol{E}-\boldsymbol{A})^k)=\text{rank}((\lambda\boldsymbol{E}-\boldsymbol{J})^k)$. 由此，我们可得求矩阵 $\boldsymbol{A}$ 的 Jordan 标准形的如下波尔曼算法：

(1) 求 $\boldsymbol{A}$ 的全部特征值；

(2) 对每个相异的特征根 λ_k 和每个 j，$(j=1, \cdots, n)$，求 $(\lambda_k\boldsymbol{E}-\boldsymbol{A})^j$ 的秩

$$r_j(\lambda_k) = \text{rank}((\lambda_k\boldsymbol{E}-\boldsymbol{A})^j)$$

在计算秩时，若发现对某个 $j*$，有 $r_{j*}(\lambda_k)=r_{j*+1}(\lambda_k)$，则

$$r_{j*+i}(\lambda_k) = r_{j*}(\lambda_k), i=1,2,\cdots n-j*, j>1;$$

(3) 对每个 λ_k，求

$$\begin{cases} b_1(\lambda_k) = n-2r_1(\lambda_k)+r_2(\lambda_k) \\ b_i(\lambda_k) = r_{i-1}(\lambda_k)-2r_i(\lambda_k)+r_{i+1}(\lambda_k) \end{cases}, i>1;$$

(4) 写出与 $\boldsymbol{A}$ 相似的标准型，它是由 b_i (λ_k) 个关于 λ_k 的 i 阶 Jordan 块的直和组成.

【例 3.3.6】 在复数域上用波尔曼算法求矩阵 $\boldsymbol{A}=\begin{pmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & -5 \end{pmatrix}$ 的 Jordan 标准形 $\boldsymbol{J}$.

解 由于 $|\lambda\boldsymbol{E}-\boldsymbol{A}|=\begin{vmatrix} \lambda-3 & 0 & -8 \\ -3 & \lambda+1 & -6 \\ 2 & 0 & \lambda+5 \end{vmatrix}=(\lambda+1)^3$，因此 $\boldsymbol{A}$ 的特征值是 $\lambda_1=\lambda_2=\lambda_3=-1$.

$$-\boldsymbol{E}-\boldsymbol{A}=\begin{pmatrix} -4 & 0 & -8 \\ -3 & 0 & -6 \\ 2 & 0 & 4 \end{pmatrix}=\begin{pmatrix} -4 \\ -3 \\ 2 \end{pmatrix}(1 \quad 0 \quad 2)$$

$$(\boldsymbol{E}-\boldsymbol{A})^2=0$$

因此 $r_1(-1)=\text{rank}(-\boldsymbol{E}-\boldsymbol{A})=1$，$r_2(-1)=\text{rank}((-\boldsymbol{E}-\boldsymbol{A})^2)=0$ 从而 $r_3(-1)=0$.

以 -1 为特征值的一阶 Jordan 块的个数是 $3-2r_1(-1)+r_2(-1)=3-2+0=1$，

以 -1 为特征值的二阶 Jordan 块的个数是 $r_1(-1)-2r_2(-1)+r_3(-1)=1-2\times0+0=1$.

因此 $\boldsymbol{A}$ 的 Jordan 标准形 $\boldsymbol{J}$ 为

$$\boldsymbol{J}=\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$

【例 3.3.7】 在复数域上求矩阵 $\boldsymbol{A}=\begin{pmatrix} -4 & 2 & 10 \\ -4 & 3 & 7 \\ -3 & 1 & 7 \end{pmatrix}$ 的 Jordan 标准形 $\boldsymbol{J}$.

解法一 由于 $\lambda \boldsymbol{E}-\boldsymbol{A}=\begin{pmatrix}\lambda+4 & -2 & -10\\ 4 & \lambda-3 & -7\\ 3 & -1 & \lambda-7\end{pmatrix}\sim\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & (\lambda-2)^3\end{pmatrix}$，因此 $\boldsymbol{A}$ 的 Jordan 标准形 $\boldsymbol{J}$ 为

$$\boldsymbol{J}=\begin{pmatrix}2 & 1 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{pmatrix}$$

解法二 $\lambda \boldsymbol{E}-\boldsymbol{A}=\begin{pmatrix}\lambda+4 & -2 & -10\\ 4 & \lambda-3 & -7\\ 3 & -1 & \lambda-7\end{pmatrix}$，$|\lambda \boldsymbol{E}-\boldsymbol{A}|=(\lambda-2)^3$

$$2\boldsymbol{E}-\boldsymbol{A}=\begin{pmatrix}6 & -2 & -10\\ 4 & -1 & -7\\ 3 & -1 & -5\end{pmatrix},\quad r_1(2)=\operatorname{rank}(2\boldsymbol{E}-\boldsymbol{A})=2$$

$$(2\boldsymbol{E}-\boldsymbol{A})^2=\begin{pmatrix}-2 & 0 & 4\\ -1 & 0 & 2\\ -1 & 0 & 2\end{pmatrix},\quad r_2(2)=\operatorname{rank}[(2\boldsymbol{E}-\boldsymbol{A})^2]=1$$

$$(2\boldsymbol{E}-\boldsymbol{A})^3=0, r_3(2)=\operatorname{rank}((2\boldsymbol{E}-\boldsymbol{A})^3)=0$$

因此 $r_4(2)=0$.

以 2 为特征值的一阶 Jordan 块的个数是 $3-2r_1(2)+r_2(2)=3-4+1=0$，

以 2 为特征值的二阶 Jordan 块的个数是 $r_1(2)-2r_2(2)+r_3(2)=2-2\times1+0=0$，

以 2 为特征值的三阶 Jordan 块的个数是 $r_2(2)-2r_3(2)+r_4(2)=1-2\times0+0=1$. 因此 $\boldsymbol{A}$ 的 Jordan 标准形 $\boldsymbol{J}$ 为

$$\boldsymbol{J}=\begin{pmatrix}2 & 1 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{pmatrix}$$

解法三 由于 $\lambda \boldsymbol{E}-\boldsymbol{A}=\begin{pmatrix}\lambda+4 & -2 & -10\\ 4 & \lambda-3 & -7\\ 3 & -1 & \lambda-7\end{pmatrix}$，$|\lambda \boldsymbol{E}-\boldsymbol{A}|=(\lambda-2)^3$，因此 $\boldsymbol{A}$ 的 Jordan 标准形只可能是如下三个矩阵中的一个，即

$$\boldsymbol{J}_1=\begin{pmatrix}2 & 1 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{pmatrix},\quad \boldsymbol{J}_2=\begin{pmatrix}2 & 1 & 0\\ 0 & 2 & 0\\ 0 & 0 & 2\end{pmatrix},\quad \boldsymbol{J}_3=\begin{pmatrix}2 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 2\end{pmatrix}$$

$$2\boldsymbol{E}-\boldsymbol{A}=\begin{pmatrix}6 & -2 & -10\\ 4 & -1 & -7\\ 3 & -1 & -5\end{pmatrix},\quad \operatorname{rank}(2\boldsymbol{E}-\boldsymbol{A})=2$$

由于 $\operatorname{rank}(2\boldsymbol{E}-\boldsymbol{J})=\operatorname{rank}(2\boldsymbol{E}-\boldsymbol{A})=2$，因此

$$\boldsymbol{J}=\begin{pmatrix}2 & 1 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{pmatrix}$$

解法四 由于

$$\lambda E-A=\begin{bmatrix}\lambda+4 & -2 & -10\\ 4 & \lambda-3 & -7\\ 3 & -1 & \lambda-7\end{bmatrix},|\lambda E-A|=(\lambda-2)^3$$

$$2E-A=\begin{bmatrix}6 & -2 & -10\\ 4 & -1 & -7\\ 3 & -1 & -5\end{bmatrix},\mathrm{rank}(2E-A)=2$$

由于 rank（$2E-J$）＝rank（$2E-A$）＝2，因此以 2 为特征值的 Jordan 块只有一个，故

$$J=\begin{bmatrix}2 & 1 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{bmatrix}$$

【例 3.3.8】 在复数域上求矩阵 $A=\begin{bmatrix}3 & 7 & -3\\ -2 & -5 & 2\\ -4 & -10 & 3\end{bmatrix}$ 的 Jordan 标准形 J.

解 由于 $|\lambda E-A|=(\lambda-1)(\lambda^2+1)$，因此 A 的初等因子是 $\lambda-1$，$\lambda-i$，$\lambda+i$，它们全是一次因子，因此 A 的 Jordan 标准形 J 为

$$J=\begin{bmatrix}1 & 0 & 0\\ 0 & i & 0\\ 0 & 0 & -i\end{bmatrix}$$

【例 3.3.9】 设 n 阶矩阵 A 满足 $A^2=0$，且 $\mathrm{rank}(A)=r$，求 A 的 Jordan 标准形 J_A.

解 因为矩阵 A 满足 $A^2=0$，所以 A 的特征值全是零，且 $J_A^2=0$，因此 J_A 的 Jordan 块至多是二阶的，即 A 的 Jordan 是二阶块或一阶块．由于 A 的二阶 Jordan 块的秩是一，A 的一阶 Jordan 块的秩是零，而 $\mathrm{rank}(A)=r$，因此 A 的 Jordan 标准是

$$J=\mathrm{diag}(J_2,\cdots,J_2,0,\cdots 0)$$

其中 $J_2=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}$ 的个数是 r，0 的个数是 $n-2r$.

以上介绍了当矩阵 A 给定后，如何求其 Jordan 标准形的方法．在定理 3.3.1 中如果 n 阶矩阵 A 与一个 Jordan 标准形 J 相似，则必然存在着一个可逆矩阵 P，使得 $P^{-1}AP=J$. 而这个矩阵 P 在讨论微分方程的求解问题时十分重要，以下我们来看看如何求这个可逆矩阵 P.

J 的某个 Jordan 块 J_i，其阶数为 m_i，主对角线上的元素为 λ_i，它们在 J 中的行数和列数分别为

$$m_1+m_2+\cdots+m_{i-1}+1,m_1+m_2+\cdots+m_{i-1}+2,\cdots,m_1+m_2+\cdots+m_{i-1}+m_i.$$

记 P 的列向量为 η_1，η_2，…，η_n，则

$$P=(\eta_1,\cdots,\eta_{m_1+m_2+\cdots+m_{i-1}+1},\cdots,\eta_{m_1+m_2+\cdots+m_{i-1}+m_i},\cdots,\eta_n)$$

因为 $P^{-1}AP=J$，所以 $AP=PJ$，

即

$$A(\eta_1,\cdots,\eta_{m_1+m_2+\cdots+m_{i-1}+1},\cdots,\eta_{m_1+m_2+\cdots+m_{i-1}+m_i},\cdots,\eta_n)$$

$$=(\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_{m_1+m_2+\cdots+m_{i-1}+1},\cdots,\boldsymbol{\eta}_{m_1+m_2+\cdots+m_{i-1}+m_i},\cdots,\boldsymbol{\eta}_n)\boldsymbol{J}$$

为方便起见，将 $m_1+m_2+\cdots+m_{i-1}+1$，$m_1+m_2+\cdots+m_{i-1}+2$，…，$m_1+m_2+\cdots+m_{i-1}+m_i$ 中的编号记为 1，2，…，m_i. 那么，

$$\boldsymbol{A\eta}_1=\lambda\boldsymbol{\eta}_1$$
$$\boldsymbol{A\eta}_2=\boldsymbol{\eta}_1+\lambda\boldsymbol{\eta}_2$$
$$\cdots$$
$$\boldsymbol{A\eta}_{m_i}=\boldsymbol{\eta}_{m_{i-1}}+\lambda\boldsymbol{\eta}_i$$

即

$$(\lambda\boldsymbol{E}-\boldsymbol{A})\boldsymbol{\eta}_1=0$$
$$(\lambda\boldsymbol{E}-\boldsymbol{A})\boldsymbol{\eta}_2=-\boldsymbol{\eta}_1$$
$$\cdots$$
$$(\lambda\boldsymbol{E}-\boldsymbol{A})\boldsymbol{\eta}_{m_i}=-\boldsymbol{\eta}_{m_{i-1}}$$

记 $\boldsymbol{H}=\lambda\boldsymbol{E}-\boldsymbol{A}$，并设 $S(\boldsymbol{H})$ 为由 $\boldsymbol{H}$ 的列向量所张成的线性空间，因为 $\boldsymbol{\eta}_1\in S(\boldsymbol{H})$ 是 $H\boldsymbol{\eta}_2=-\boldsymbol{\eta}_1$ 有解的充要条件，而 $\boldsymbol{\eta}_1\in S(\boldsymbol{H})$ 的充要条件是 $\boldsymbol{\eta}_1$ 与 $S(\boldsymbol{H})^{\perp}$ 中的一切向量正交，也就是与 $S(\boldsymbol{H})^{\perp}$ 的一个基正交．同时，$S(\boldsymbol{H})^{\perp}$ 的一个基即为方程 $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{X}=0$ 的一个基础解系，为使 $H\boldsymbol{\eta}_2=-\boldsymbol{\eta}_1$ 对 $\boldsymbol{\eta}_2$ 有解，求 $\boldsymbol{\eta}_1$ 方法如下：

(1) 求方程 $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{X}=\boldsymbol{0}$ 的一个基础解系 $\boldsymbol{p}_1$，$\boldsymbol{p}_2$，…，$\boldsymbol{p}_s$，并以 $\boldsymbol{p}_1^{\mathrm{T}}$，$\boldsymbol{p}_2^{\mathrm{T}}$，…，$\boldsymbol{p}_s^{\mathrm{T}}$ 作为行向量构成矩阵 $\boldsymbol{B}$；

(2) 以方程组 $\begin{pmatrix}\boldsymbol{H}\\\boldsymbol{B}\end{pmatrix}\boldsymbol{X}=\boldsymbol{0}$ 的一个非零解作为 $\boldsymbol{\eta}_1$；

(3) 同理得到 $\boldsymbol{\eta}_k(k=1,2,\cdots,m_{i-1})$，$\boldsymbol{\eta}_k$ 作为 $\begin{pmatrix}\boldsymbol{H}\\\boldsymbol{B}\end{pmatrix}\boldsymbol{X}=\begin{pmatrix}-\boldsymbol{\eta}_{k-1}\\0\end{pmatrix}$ 的一个非零解；

(4) $\boldsymbol{\eta}_{m_i}$ 为 $\boldsymbol{HX}=-\boldsymbol{\eta}_{m_{i-1}}$ 的非零解．

由此解出 $\boldsymbol{P}=(\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_{m_1+m_2+\cdots+m_{i-1}+1},\cdots,\boldsymbol{\eta}_{m_1+m_2+\cdots+m_{i-1}+m_i},\cdots,\boldsymbol{\eta}_n)$.

【例 3.3.10】 求矩阵 $\boldsymbol{A}=\begin{pmatrix}-4&2&10\\-4&3&7\\-3&1&7\end{pmatrix}$ 的 Jordan 标准形 $\boldsymbol{J}$，并求出可逆矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{J}$.

解 由初等变换可得 $\boldsymbol{A}=\begin{pmatrix}-4&2&10\\-4&3&7\\-3&1&7\end{pmatrix}\to\begin{pmatrix}1&0&0\\0&1&0\\0&0&(\lambda-2)^3\end{pmatrix}$，所以，$\boldsymbol{A}$ 与 Jordan 标准形 $\boldsymbol{J}=\begin{pmatrix}2&1&0\\0&2&1\\0&0&2\end{pmatrix}$ 相似．

令 $\boldsymbol{H}=2\boldsymbol{E}-\boldsymbol{A}=\begin{pmatrix}6&-2&-10\\4&-1&-7\\3&-1&-5\end{pmatrix}$，求可逆矩阵步骤如下：

(1) 求解方程组 $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{X}=\boldsymbol{0}$，得到 $\boldsymbol{X}=k\begin{pmatrix}1\\0\\-2\end{pmatrix}$，取 $\boldsymbol{B}=(1,0,-2)$；

（2）由 $\begin{pmatrix}\boldsymbol{H}\\ \boldsymbol{B}\end{pmatrix}\boldsymbol{X}=0$，得到 $\boldsymbol{X}=k\begin{pmatrix}2\\1\\1\end{pmatrix}$，取 $\boldsymbol{\eta}_1=\begin{pmatrix}2\\1\\1\end{pmatrix}$；

（3）由 $\begin{pmatrix}\boldsymbol{H}\\ \boldsymbol{B}\end{pmatrix}\boldsymbol{X}=\begin{pmatrix}-\boldsymbol{\eta}_1\\ 0\end{pmatrix}$，得到 $\boldsymbol{X}=k\begin{pmatrix}0\\1\\0\end{pmatrix}$，取 $\boldsymbol{\eta}_2=\begin{pmatrix}0\\1\\0\end{pmatrix}$；

（4）由 $\boldsymbol{HX}=-\boldsymbol{\eta}_2$，得到 $\boldsymbol{X}=k\begin{pmatrix}-1\\-3\\0\end{pmatrix}$，取 $\boldsymbol{\eta}_3=\begin{pmatrix}-1\\-3\\0\end{pmatrix}$.

所以，$\boldsymbol{P}=(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\eta}_3)=\begin{pmatrix}2&0&-1\\1&1&-3\\1&0&0\end{pmatrix}$. 检验有 $\boldsymbol{AP}=\boldsymbol{PJ}=\begin{pmatrix}4&2&-2\\2&3&-5\\2&1&0\end{pmatrix}$，即 $\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{J}$.

前面，我们已经介绍，如果 n 阶矩阵 $\boldsymbol{A}$ 与一个 Jordan 标准形 $\boldsymbol{J}$ 相似，则必然存在着一个可逆矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{J}$. 如果将 $\boldsymbol{P}$ 按列分块成 $\boldsymbol{P}=(\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_n)$，则有

$$(\boldsymbol{Ap}_1,\boldsymbol{Ap}_2,\cdots,\boldsymbol{Ap}_n)=(\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_n)\boldsymbol{J}$$

由此可得如下形式的齐次或非齐次线性方程组

$$\boldsymbol{Ap}_i=a\boldsymbol{p}_{i-1}+\lambda_i\boldsymbol{p}_i,\quad i=1,2,\cdots,n$$

式中：$a=0$ 或 1. 由这些方程组可求得 $\boldsymbol{P}=(\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_n)$.

【例 3.3.11】 求矩阵 $\boldsymbol{A}=\begin{pmatrix}-4&2&10\\-4&3&7\\-3&1&7\end{pmatrix}$ 的 Jordan 标准形 $\boldsymbol{J}$，并求出可逆矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{J}$.

解 由前例有，$\boldsymbol{A}$ 与 Jordan 标准形 $\boldsymbol{J}=\begin{pmatrix}2&1&0\\0&2&1\\0&0&2\end{pmatrix}$ 相似. 令 $\boldsymbol{P}=(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3)$，则有

$$\boldsymbol{Ap}_1=2\boldsymbol{p}_1,\boldsymbol{Ap}_2=\boldsymbol{p}_1+2\boldsymbol{p}_2,\boldsymbol{Ap}_3=\boldsymbol{p}_2+2\boldsymbol{p}_3$$

即

$$\begin{pmatrix}-6&2&10\\-4&1&7\\-3&1&5\end{pmatrix}\boldsymbol{p}_1=\begin{pmatrix}0\\0\\0\end{pmatrix},\begin{pmatrix}-6&2&10\\-4&1&7\\-3&1&5\end{pmatrix}\boldsymbol{p}_2=\boldsymbol{p}_1,\begin{pmatrix}-6&2&10\\-4&1&7\\-3&1&5\end{pmatrix}\boldsymbol{p}_3=\boldsymbol{p}_2$$

解得 $\boldsymbol{p}_1=(2,1,1)^{\mathrm{T}}$，$\boldsymbol{p}_2=(0, 1, 0)^{\mathrm{T}}$，$\boldsymbol{p}_3=(1,-2,1)^{\mathrm{T}}$，故

$$\boldsymbol{P}=\begin{pmatrix}2&0&1\\1&1&-2\\1&0&1\end{pmatrix}$$

【例 3.3.12】 求矩阵 $\boldsymbol{A}=\begin{pmatrix}3&1&-1\\-2&0&2\\-1&-1&3\end{pmatrix}$ 的 Jordan 标准形 $\boldsymbol{J}$，并求出可逆矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{J}$.

解 $|\lambda E - A| = (\lambda-2)^3$，

$$A-2E=\begin{pmatrix}1&1&-1\\-2&-2&2\\-1&-1&1\end{pmatrix}\sim\begin{pmatrix}1&1&-1\\0&0&0\\0&0&0\end{pmatrix}$$

因此 Jordan 标准形

$$J=\begin{pmatrix}2&0&0\\0&2&1\\0&0&2\end{pmatrix}$$

A 的对应于$\lambda=2$ 的线性无关特征向量为 $\xi_1=(1,-1,0)^T$，$\xi_2=(1,0,1)^T$. 设可逆矩阵 $P=(p_1,p_2,p_3)$，使得 $P^{-1}AP=J$，则

$$Ap_1=2p_1, Ap_2=2p_2, Ap_3=p_2+2p_3$$

可见，p_1，p_2 是 A 的对应于$\lambda=2$ 的线性无关特征向量．如果直接取，$p_1=\xi_1$，$p_2=\xi_2$，则

$$(A-2E,p_2)=\begin{pmatrix}1&1&-1&1\\-2&-2&2&0\\-1&-1&1&1\end{pmatrix}\sim\begin{pmatrix}1&1&-1&1\\0&0&0&1\\0&0&0&0\end{pmatrix}$$

则方程 $(A-2E)x=p_2$ 无解．注意到对应于同一特征值的特征向量的非零组合仍然是特征向量，为使方程 $(A-2E)x=p_2$ 有解，取 $p_2=k_1\xi_1+k_2\xi_2$

$$(A-2E,p_2)=\begin{pmatrix}1&1&-1&k_1+k_2\\-2&-2&2&-k_1\\-1&-1&1&k_2\end{pmatrix}\sim\begin{pmatrix}1&1&-1&1\\0&0&0&k_1+2k_2\\0&0&0&0\end{pmatrix}$$

因此，方程 $(A-2E)x=p_2$ 有解的充分必要条件是 $k_1=-2k_2$. 特别取 $p_2=-2\xi_1+\xi_2=(-1,2,1)^T$，此时，可解得 $p_3=(1,0,0)^T$，因此

$$P=\begin{pmatrix}1&-1&1\\-1&2&0\\0&1&0\end{pmatrix}$$

在后面的章节中，我们常用矩阵的 Jordan 标准形进行一些理论推证、简化矩阵的计算，如求方幂等．此时要用到求 Jordan 块的方幂的如下结果

$$J_i=\begin{pmatrix}\lambda_i&1&&&\\&\lambda_i&1&&\\&&\lambda_i&\ddots&\\&&&\ddots&1\\&&&&\lambda_i\end{pmatrix}_{m\times m},J^k=\begin{pmatrix}\lambda^k&C_k^1\lambda^k&C_k^2\lambda^k&\cdots&C_k^{m-1}\lambda^k\\0&\lambda^k&C_k^1\lambda^{k-1}&\cdots&C_k^{m-2}\lambda^k\\0&0&\lambda^k&\cdots&\vdots\\\vdots&\vdots&\vdots&\ddots&C_k^1\lambda^{k-1}\\0&0&0&\cdots&\lambda^k\end{pmatrix}_{m\times m}$$

其中 $C_k^r=\dfrac{k(k-1)\cdots(k-r+1)}{r!}$，且当 $r>k$ 时，规定 $C_k^r=0$.

【例 3.3.13】 设数列 $\{a_n\}$ 满足条件

$$a_n=3a_{n-2}+2a_{n-3}, \forall n\geqslant 3$$

已知 a_1，a_2，a_3，求 a_n.

解
$$\begin{pmatrix}a_n\\a_{n-1}\\a_{n-2}\end{pmatrix}=\begin{pmatrix}3a_{n-2}+2a_{n-3}\\a_{n-1}\\a_{n-2}\end{pmatrix}=\begin{pmatrix}0&3&2\\1&0&0\\0&1&0\end{pmatrix}\begin{pmatrix}a_{n-1}\\a_{n-2}\\a_{n-3}\end{pmatrix}=\begin{pmatrix}0&3&2\\1&0&0\\0&1&0\end{pmatrix}^{n-3}\begin{pmatrix}a_3\\a_2\\a_1\end{pmatrix}$$

由初等变换可得

$$\lambda \boldsymbol{E}-\boldsymbol{A}=\begin{pmatrix}\lambda & -3 & -2\\ -1 & \lambda & 0\\ 0 & -1 & \lambda\end{pmatrix}\rightarrow\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & (\lambda-2)(\lambda+1)^2\end{pmatrix}$$

所以，$\boldsymbol{A}$ 与 Jordan 标准形 $\boldsymbol{J}=\begin{pmatrix}2 & 0 & 0\\ 0 & -1 & 1\\ 0 & 0 & -1\end{pmatrix}$ 相似．

$$\boldsymbol{A}-2\boldsymbol{E}=\begin{pmatrix}-2 & 3 & 2\\ 1 & -2 & 0\\ 0 & 1 & -2\end{pmatrix}\sim\begin{pmatrix}1 & 0 & -4\\ 0 & 1 & -2\\ 0 & 0 & 0\end{pmatrix}.$$

$\boldsymbol{A}$ 的对应于 $\lambda=2$ 的线性无关特征向量为 $\boldsymbol{p}_1=(4,2,1)^{\mathrm{T}}$．

$$\boldsymbol{A}+\boldsymbol{E}=\begin{pmatrix}1 & 3 & 2\\ 1 & 1 & 0\\ 0 & 1 & 1\end{pmatrix}\sim\begin{pmatrix}1 & 0 & -1\\ 0 & 1 & 1\\ 0 & 0 & 0\end{pmatrix}$$

$\boldsymbol{A}$ 的对应于 $\lambda_2=-1$ 的线性无关特征向量为 $\boldsymbol{p}_2=(1,-1,1)^{\mathrm{T}}$．

$$(\boldsymbol{A}+\boldsymbol{E},\boldsymbol{p}_2)=\begin{pmatrix}1 & 3 & 2 & 1\\ 1 & 1 & 0 & -1\\ 0 & 1 & 1 & 1\end{pmatrix}\sim\begin{pmatrix}1 & 0 & -1 & -2\\ 0 & 1 & 1 & 1\\ 0 & 0 & 0 & 0\end{pmatrix}$$

解得 $\boldsymbol{p}_3=(-2,1,0)^{\mathrm{T}}$．因此，取

$$\boldsymbol{P}=\begin{pmatrix}4 & 1 & -2\\ 2 & -1 & 1\\ 1 & 1 & 0\end{pmatrix}$$

则 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}$.

$$\boldsymbol{A}^{n-3}=\boldsymbol{P}\begin{pmatrix}2 & 0 & 0\\ 0 & -1 & 1\\ 0 & 0 & -1\end{pmatrix}^{n-3}\boldsymbol{P}^{-1}$$

$$=\frac{1}{9}\begin{pmatrix}4 & 1 & -2\\ 2 & -1 & 1\\ 1 & 1 & 0\end{pmatrix}\begin{pmatrix}2^{n-3} & 0 & 0\\ 0 & (-1)^{n-3} & (-1)^{n-2}(n-3)\\ 0 & 0 & (-1)^{n-3}\end{pmatrix}\begin{pmatrix}1 & 2 & 1\\ -1 & -2 & 8\\ -3 & 3 & 6\end{pmatrix}$$

$$=\frac{1}{9}\begin{pmatrix}2^{n-1}+(-1)^{n-1}(3n-4) & 2^{n-1}+(-1)^{n}(3n-1) & 2^{n-1}+(-1)^{n}(6n-14)\\ * & * & *\\ * & * & *\end{pmatrix}$$

因此

$$a_n=\frac{2^{n-1}+(-1)^{n-1}(3n-4)}{9}a_3+\frac{2^{n-1}+(-1)^{n}(3n-1)}{9}a_2+\frac{2^{n-1}+(-1)^{n}(6n-14)}{9}a_1$$

应用矩阵的 Jordan 标准形还可用于求解线性微分方程组．

对于线性微分方程组，其矩阵形式为 $\frac{\mathrm{d}x}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}$. 如果 $\boldsymbol{A}$ 的 Jordan 标准形是 $\boldsymbol{J}$，而可逆矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}$，则 $\frac{\mathrm{d}x}{\mathrm{d}t}=\boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1}\boldsymbol{x}$．令，$\boldsymbol{y}=\boldsymbol{P}^{-1}\boldsymbol{x}$，则有 $\frac{\mathrm{d}y}{\mathrm{d}t}=\boldsymbol{J}\boldsymbol{y}$，它们是一系列容易

求解的一阶线性齐次（非齐次）方程．求出 $\boldsymbol{y}$ 后，则有 $\boldsymbol{x}=\boldsymbol{P}\boldsymbol{y}$，从而求得问题的解．

【例 3.3.14】 解如下线性微分方程组

$$\begin{cases}\dfrac{\mathrm{d}x_1}{\mathrm{d}t}=3x_1+x_2-x_3\\ \dfrac{\mathrm{d}x_2}{\mathrm{d}t}=-2x_1+2x_3\\ \dfrac{\mathrm{d}x_3}{\mathrm{d}t}=-x_1-x_2+3x_3\end{cases}$$

解 记 $\boldsymbol{x}=(x_1(t),x_2(t),x_3(t))^{\mathrm{T}}$，$\boldsymbol{A}=\begin{pmatrix}3&1&-1\\-2&0&2\\-1&-1&3\end{pmatrix}$，则线性微分方程组的矩阵形式为

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}$$

由例 3.3.12 知，矩阵 $\boldsymbol{P}=\begin{pmatrix}1&-1&1\\-1&2&0\\0&1&0\end{pmatrix}$ 使得

$$\boldsymbol{A}=\boldsymbol{P}\begin{pmatrix}2&0&0\\0&2&1\\0&0&1\end{pmatrix}\boldsymbol{P}^{-1}$$

令 $\boldsymbol{y}=\boldsymbol{P}^{-1}\boldsymbol{x}$，则

$$\begin{cases}\dfrac{\mathrm{d}y_1}{\mathrm{d}t}=2y_1\\ \dfrac{\mathrm{d}y_2}{\mathrm{d}t}=2y_2+y_3\\ \dfrac{\mathrm{d}y_3}{\mathrm{d}t}=2y_3\end{cases}$$

因此 $y_1=c_1e^{2t}$，$y_2=(c_2+c_3t)e^{2t}$，$y_3=c_3e^{2t}$，从而原线性微分方程组的解为

$$\boldsymbol{x}=\begin{pmatrix}1&-1&1\\-1&2&0\\0&1&0\end{pmatrix}\begin{pmatrix}c_1e^{2t}\\(c_2+c_3t)e^{2t}\\c_3e^{2t}\end{pmatrix}=\begin{pmatrix}c_1-c_2-c_3t+c_3\\-c_1+2c_2+2c_3t\\c_2+c_3t\end{pmatrix}e^{2t}$$

其中 c_1，c_2，c_3 为任意常数．

【例 3.3.15】 设 $\boldsymbol{\xi}_1$，$\boldsymbol{\xi}_2$，$\boldsymbol{\xi}_3$，$\boldsymbol{\xi}_4$ 是线性空间 $\boldsymbol{V}$ 的一组基，$\boldsymbol{V}$ 上的线性变换 $\boldsymbol{T}$ 在该基下的矩阵

$$\boldsymbol{A}=\begin{pmatrix}6&5&-8&6\\-2&-1&4&-2\\1&1&0&2\\0&0&0&2\end{pmatrix}$$

求线性变换 $\boldsymbol{T}$ 的所有不变子空间．

解 $|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-1)(\lambda-2)^3$，

$$A-2E=\begin{pmatrix}4&5&-8&6\\-2&-3&4&-2\\1&1&-2&2\\0&0&0&0\end{pmatrix}\sim\begin{pmatrix}0&1&0&-2\\0&-1&0&2\\1&1&-2&2\\0&0&0&0\end{pmatrix}\sim\begin{pmatrix}1&1&-2&2\\0&1&0&-2\\0&0&0&0\\0&0&0&0\end{pmatrix}$$

因此 A 的 Jordan 标准形 $J=\begin{pmatrix}1&0&0&0\\0&2&0&0\\0&0&2&1\\0&0&0&2\end{pmatrix}$.

$$A-2E\sim\begin{pmatrix}1&0&-2&4\\0&1&0&-2\\0&0&0&0\\0&0&0&0\end{pmatrix}$$

A 的对应于 $\lambda=2$ 的线性无关特征向量为 $\xi_2=(2,0,1,0)^T$，$\xi_3=(-4,2,0,1)^T$

设可逆矩阵 $P=(p_1,p_2,p_3,p_4)$，使得 $P^{-1}AP=J$，则有

$$Ap_1=p_1, Ap_2=2p_2, Ap_3=2p_3, Ap_4=p_3+2p_4$$

$$A-E=\begin{pmatrix}5&5&-8&6\\-2&-2&4&-2\\1&1&-1&2\\0&0&0&1\end{pmatrix}\sim\begin{pmatrix}1&1&0&0\\0&0&1&0\\0&0&0&1\\0&0&0&0\end{pmatrix}$$

A 的对应于 $\lambda=1$ 的线性无关特征向量为 $p_1=(1,-1,0,0)^T$.

取 $p_3=k_1(2,0,1,0)^T+k_2(-4,2,0,1)^T$

$$A-2E=\begin{pmatrix}4&5&-8&6\\-2&-3&4&-2\\1&1&-2&2\\0&0&0&0\end{pmatrix}\sim\begin{pmatrix}0&1&0&-2\\0&-1&0&2\\1&1&-2&2\\0&0&0&0\end{pmatrix}\sim\begin{pmatrix}1&1&-2&2\\0&1&0&-2\\0&0&0&0\\0&0&0&0\end{pmatrix}$$

$$(A-2E,p_3)=\begin{pmatrix}4&5&-8&6&2k_1-4k_2\\-2&-3&4&-2&2k_2\\1&1&-2&2&k_1\\0&0&0&0&k_2\end{pmatrix}$$

因此可取 $p_2=(-4,2,0,1)^T$，$p_3=(2,0,1,0)^T$，

$$(A-2E,p_3)=\begin{pmatrix}4&5&-8&6&2\\-2&-3&4&-2&0\\1&1&-2&2&1\\0&0&0&0&0\end{pmatrix}\sim\begin{pmatrix}1&0&-2&4&3\\0&1&0&-2&-2\\0&0&0&0&0\\0&0&0&0&0\end{pmatrix}$$

$p_4=(5,-2,1,0)^T$，因此

$$P=\begin{pmatrix}1&-4&2&5\\-1&2&0&-2\\0&0&1&1\\0&0&1&0\end{pmatrix}$$

令

$$e_1=\xi_1-\xi_2, e_2=-4\xi_1+2\xi_2, e_3=2\xi_1+\xi_3+\xi_4, e_4=5\xi_1-2\xi_2+\xi_3$$

$$V_1=\text{span}\{e_1\}, V_2=\text{span}\{e_2\}, V_3=\text{span}\{e_3, e_3+2e_4\},$$

则线性变换 T 的不变子空间有

$$V_0=\{0\}, V_1, V_2, V_3, V_1\oplus V_2, V_1\oplus V_3, V_2\oplus V_3, V$$

在 Matlab 中，求 A 的 Jordan 标准形的命令是 J=Jordan（A）. 如果要求相应的变换阵 P，则命令为 $[P,J]=\text{Jordan}(A)$

【例 3.3.16】 求矩阵 $A=\begin{pmatrix}1&2&5&4\\2&-1&2&0\\-1&3&1&1\\3&4&-3&4\end{pmatrix}$ 的 Jordan 标准形 J，并求出可逆矩阵 P，使得 $P^{-1}AP=J$.

解 ≫A=[1 2 5 4;2 −1 2 0;−1 3 1 1;3 4 −3 4]；

≫[P,J]=Jordan(A)

其运算结果为

P=

−0.0560	0.3229	0.3666−0.0571i	0.3666+0.0571i
−0.0960	0.0931	0.0014−0.2713i	0.0014+0.2713i
−0.0880	0.0548	0.0166+0.2023i	0.0166−0.2023i
0.1440	0.3782	−0.2611+0.1857i	−0.2611−0.1857i

J=

2.0000	0	0	0
0	7.1111	0	0
0	0	−2.0555+2.8300i	0
0	0	0	−2.0555−2.8300i

§3.4 Cayley-Hamilton 定理 最小多项式

根据多项式的形式表示法，可以给出 λ-矩阵的一个记法.

例如，矩阵 $A(\lambda)=\begin{bmatrix}\lambda^4+\lambda & \lambda^2-\lambda+1\\ \lambda^3+1 & \lambda^2+\lambda+2\end{bmatrix}$，

可以记为

$$A(\lambda)=\begin{bmatrix}\lambda^4+0\lambda^3+0\lambda^2+\lambda+0 & 0\lambda^4+0\lambda^3+\lambda^2-\lambda+1\\ 0\lambda^4+\lambda^3+0\lambda^2+0\lambda+1 & 0\lambda^4+0\lambda^3+\lambda^2+\lambda+2\end{bmatrix}$$

$$=\begin{pmatrix}1&0\\0&0\end{pmatrix}\lambda^4+\begin{pmatrix}0&0\\1&0\end{pmatrix}\lambda^3+\begin{pmatrix}0&1\\0&1\end{pmatrix}\lambda^2+\begin{pmatrix}1&-1\\0&1\end{pmatrix}\lambda+\begin{pmatrix}0&1\\1&2\end{pmatrix}.$$

一般的，一个次数不大于 m 的 n 阶 λ-矩阵 $B(\lambda)$ 可以记为

$$B(\lambda)=B_0\lambda^m+B_1\lambda^{m-1}+\cdots+B_{m-1}\lambda+B_m$$

此式叫做 n 阶 λ-矩阵 $B(\lambda)$ 的多项式记法，其中 B_i（i=0，1，…，m）为 n 阶常数矩阵.

形如 $\varphi(\boldsymbol{A})=a_0\boldsymbol{A}^m+a_1\boldsymbol{A}^{m-1}+\cdots+a_{m-1}\boldsymbol{A}+a_m\boldsymbol{E}$ 的式子叫做矩阵 $\boldsymbol{A}$ 的多项式，其中 $\boldsymbol{A}\in \mathbf{C}^{n\times n}$，$a_i$ 是常数，$m\in\mathbf{Z}^+$. 当然它也是一个 n 阶矩阵.

定理 3.4.1　(Cayley-Hamilton 定理)　设 $f(\lambda)=\det(\lambda\boldsymbol{E}-\boldsymbol{A})$ 为矩阵 $\boldsymbol{A}$ 的特征多项式，则 $f(\boldsymbol{A})=0$.

证法一　设 $\boldsymbol{J}=\begin{pmatrix}\boldsymbol{J}_1 & & & \\ & \boldsymbol{J}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_s\end{pmatrix}$ 是矩阵 $\boldsymbol{A}$ 的 Jordan 标准形，其中 $\boldsymbol{J}_i$ ($i=1,2,\cdots,s$) 为 m_i 阶 Jordan 块，可逆矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{J}$. 则

$$f(\lambda)=\det(\lambda\boldsymbol{E}-\boldsymbol{A})=\prod_{k=1}^{s}(\lambda-\lambda_k)^{m_k}$$

$$f(\boldsymbol{A})=\prod_{k=1}^{s}(\boldsymbol{A}-\lambda_k\boldsymbol{E})^{m_k}=\boldsymbol{P}\prod_{k=1}^{s}(\boldsymbol{J}-\lambda_k\boldsymbol{E})^{m_k}\boldsymbol{P}^{-1}$$

注意到

$$(\boldsymbol{J}-\lambda_k\boldsymbol{E})^{m_k}=\begin{pmatrix}(\boldsymbol{J}_1-\lambda_k\boldsymbol{E})^{m_k} & & & \\ & (\boldsymbol{J}_2-\lambda_k\boldsymbol{E})^{m_k} & & \\ & & \ddots & \\ & & & (\boldsymbol{J}_s-\lambda_k\boldsymbol{E})^{m_k}\end{pmatrix}$$

注意到 $(\boldsymbol{J}_k-\lambda_k\boldsymbol{E})^{m_k}=\begin{pmatrix}0&1&0&\cdots&0\\0&0&1&\cdots&0\\0&0&0&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&0&0\end{pmatrix}_{m_k\times m_k}^{m_k}=0$，因此

$$f(\boldsymbol{A})=\boldsymbol{P}\begin{pmatrix}0 & & \\ & * & \\ & & \ddots \\ & & & *\end{pmatrix}\begin{pmatrix}* & & \\ & 0 & \\ & & \ddots \\ & & & *\end{pmatrix}\cdots\begin{pmatrix}* & & \\ & * & \\ & & \ddots \\ & & & 0\end{pmatrix}\boldsymbol{P}^{-1}=0$$

证法二　设 $\boldsymbol{B}$ 为 $\lambda\boldsymbol{E}-\boldsymbol{A}$ 的伴随矩阵，则矩阵 $\boldsymbol{B}$ 中的元素均为 $\lambda\boldsymbol{E}-\boldsymbol{A}$ 中各元素的代数余子式，即为 λ 的次数不超过 $n-1$ 的多项式，所以 $\boldsymbol{B}$ 为 $n-1$ 阶的 λ-矩阵. 记作

$$\boldsymbol{B}=\boldsymbol{B}_0\lambda^{n-1}+\boldsymbol{B}_1\lambda^{n-2}+\cdots+\boldsymbol{B}_{n-2}\lambda+\boldsymbol{B}_{n-1}$$

并且由伴随矩阵的性质有

$$\boldsymbol{B}(\lambda\boldsymbol{E}-\boldsymbol{A})=|\lambda\boldsymbol{E}-\boldsymbol{A}|\boldsymbol{E}=f(\lambda)\boldsymbol{E}$$

即

$$(\boldsymbol{B}_0\lambda^{n-1}+\boldsymbol{B}_1\lambda^{n-2}+\cdots+\boldsymbol{B}_{n-2}\lambda+\boldsymbol{B}_{n-1})(\lambda\boldsymbol{E}-\boldsymbol{A})=(\lambda^n+a_1\lambda^{n-1}+\cdots+a_{n-1}\lambda+a_n)\boldsymbol{E}$$

比较等式两边的系数有

$$-\boldsymbol{B}_{n-1}\boldsymbol{A}=a_n\boldsymbol{E}$$
$$-\boldsymbol{B}_{n-2}\boldsymbol{A}+\boldsymbol{B}_{n-1}=a_{n-1}\boldsymbol{E}$$
$$-\boldsymbol{B}_{n-3}\boldsymbol{A}+\boldsymbol{B}_{n-2}=a_{n-2}\boldsymbol{E}$$
$$\cdots$$

$$-\boldsymbol{B}_0\boldsymbol{A}+\boldsymbol{B}_1=a_1\boldsymbol{E}$$
$$\boldsymbol{B}_0=\boldsymbol{E}$$

依次以 $\boldsymbol{E}$，$\boldsymbol{A}$，$\boldsymbol{A}^2$，…，$\boldsymbol{A}^n$ 右乘以上各式，两边相加得

$$0=a_n\boldsymbol{E}+a_{n-1}\boldsymbol{A}+\cdots+a_1\boldsymbol{A}^{n-1}+\boldsymbol{A}^n,\text{即 } \boldsymbol{f}(\boldsymbol{A})=0$$

【例 3.4.1】 设矩阵 $\boldsymbol{A}=\begin{pmatrix}-1&1&0\\-4&3&0\\1&0&2\end{pmatrix}$，试计算 $\boldsymbol{h}(\boldsymbol{A})=\boldsymbol{A}^7-\boldsymbol{A}^5-19\boldsymbol{A}^4+28\boldsymbol{A}^3+6\boldsymbol{A}-4\boldsymbol{E}$.

解 因为 $\boldsymbol{A}$ 的特征多项式 $\det(\lambda\boldsymbol{E}-\boldsymbol{A})=f(\lambda)=\lambda^3-4\lambda^2+5\lambda-2$，而 $h(\lambda)=\lambda^7-\lambda^5-19\lambda^4+28\lambda^3+6\lambda-4$，以 $f(\lambda)$ 除 $h(\lambda)$ 所得余式为 $-3\lambda^2+22\lambda-8$，因此，$h(\lambda)=s(\lambda)f(\lambda)+(-3\lambda^2+22\lambda-8)$，所以

$$\boldsymbol{h}(\boldsymbol{A})=-3\boldsymbol{A}^2+22\boldsymbol{A}-8\boldsymbol{E}$$
$$=-3\begin{pmatrix}-1&1&0\\-4&3&0\\1&0&2\end{pmatrix}^2+22\begin{pmatrix}-1&1&0\\-4&3&0\\1&0&2\end{pmatrix}+\begin{pmatrix}-8&0&0\\0&-8&0\\0&0&-8\end{pmatrix}$$
$$=\begin{pmatrix}-19&16&0\\-64&43&0\\19&-3&24\end{pmatrix}$$

【例 3.4.2】 已知 $\boldsymbol{A}=\begin{pmatrix}1&0&-1\\0&\omega&\sqrt{2}i\\0&0&\omega^2\end{pmatrix}$，其中 $\omega=\dfrac{-1+\sqrt{3}i}{2}$，试计算 $\boldsymbol{A}^4$.

解 $\because \det(\lambda\boldsymbol{E}-\boldsymbol{A})=f(\lambda)=(\lambda-1)(\lambda-\omega)(\lambda-\omega^2)=\lambda^3-1$

$\therefore \boldsymbol{f}(\boldsymbol{A})=\boldsymbol{A}^3-\boldsymbol{E}=0$，即 $\boldsymbol{A}^3=\boldsymbol{E}$，所以 $\boldsymbol{A}^4=\boldsymbol{A}^3\boldsymbol{A}=\boldsymbol{E}\boldsymbol{A}=\boldsymbol{A}$.

Cayley-Hamilton 定理说明，对于任意一个 n 阶矩阵 $\boldsymbol{A}$，一定存在着多项式 $\boldsymbol{\varphi}(\lambda)$，使得 $\boldsymbol{\varphi}(\boldsymbol{A})=\boldsymbol{0}$.

定义 3.4.1 凡是能使 $\boldsymbol{\varphi}(\boldsymbol{A})=\boldsymbol{0}$ 的多项式 $\varphi(\lambda)$ 叫做矩阵 $\boldsymbol{A}$ 的零化多项式.

显然，矩阵 $\boldsymbol{A}$ 的特征多项式 $f(\lambda)=\det(\lambda\boldsymbol{E}-\boldsymbol{A})$ 是 $\boldsymbol{A}$ 的一个零化多项式.

如果 $\varphi(\lambda)=0$，则对任意的多项式 $g(\lambda)$ 而言，均有 $f(\lambda)=g(\lambda)\varphi(\lambda)=0$. 因此，零化多项式不唯一.

同时，$\boldsymbol{A}$ 的特征多项式也不一定是 $\boldsymbol{A}$ 的次数最低的零化多项式.

例如，$\boldsymbol{A}=\begin{pmatrix}5&0&0\\0&3&-2\\0&-2&3\end{pmatrix}$，$f(\lambda)=\det(\lambda\boldsymbol{E}-\boldsymbol{A})=(\lambda-1)(\lambda-5)^2$，则 $\boldsymbol{f}(\boldsymbol{A})=0$，所以，$\boldsymbol{f}(\boldsymbol{A})$ 是 $\boldsymbol{A}$ 的一个零化多项式；但是，若 $h(\lambda)=(\lambda-1)(\lambda-5)$，则 $\boldsymbol{h}(\boldsymbol{A})=0$，所以，$\boldsymbol{h}(\boldsymbol{A})$ 也是 $\boldsymbol{A}$ 的一个零化多项式.

定义 3.4.2 矩阵 $\boldsymbol{A}$ 的所有零化多项式中，次数最低、且首项系数是 1 的零化多项式，叫做 $\boldsymbol{A}$ 的最小多项式，记作 $m(\lambda)$.

定理 3.4.2 多项式 $\varphi(\lambda)$ 是矩阵 $\boldsymbol{A}$ 的零化多项式的充分必要条件是 $m(\lambda)$ 整除 $\varphi(\lambda)$.

证 设 $\varphi(\lambda)$ 是矩阵 $\boldsymbol{A}$ 的零化多项式，所以，$\varphi(\lambda)$ 的次数必定大于或等于 $m(\lambda)$ 的次数，

用 $m(\lambda)$ 除 $\varphi(\lambda)$，有 $\varphi(\lambda)=m(\lambda)q(\lambda)+r(\lambda)$，其中余式 $r(\lambda)$ 或者恒等于零，或者它的次数小于 $m(\lambda)$ 的次数，则

$$\varphi(\boldsymbol{A})=m(\boldsymbol{A})q(\boldsymbol{A})+\boldsymbol{r}(\boldsymbol{A})$$

又因为 $\boldsymbol{\varphi}(\boldsymbol{A})=\boldsymbol{0}$，$\boldsymbol{m}(\boldsymbol{A})=\boldsymbol{0}$，所以 $\boldsymbol{r}(\boldsymbol{A})=\boldsymbol{0}$. 即 $r(\lambda)\equiv 0$，也就是 $m(\lambda)$ 整除 $\varphi(\lambda)$.

反之，如果 $m(\lambda)$ 整除 $\varphi(\lambda)$，则 $\varphi(\lambda)=m(\lambda)q(\lambda)$，即 $\boldsymbol{\varphi}(\boldsymbol{A})=\boldsymbol{m}(\boldsymbol{A})\boldsymbol{q}(\boldsymbol{A})=\boldsymbol{0}$，所以，$\varphi(\lambda)$ 是矩阵 $\boldsymbol{A}$ 的零化多项式.

由 Cayley-Hamilton 定理和定理 3.4.2，有下面推论.

推论 3.4.1 $\boldsymbol{A}$ 的最小多项式 $m(\lambda)$ 的零点一定是 $\boldsymbol{A}$ 的特征值.

定理 3.4.3 矩阵 $\boldsymbol{A}$ 的特征值必定是它的最小多项式的零点.

证 用反证法. 若 λ_0 是矩阵 $\boldsymbol{A}$ 的特征值，而不是 $\boldsymbol{A}$ 的最小多项式的根，则 $m(\lambda)$ 与 $\lambda-\lambda_0$ 互素，因而存在多项式 $u(\lambda)$，$v(\lambda)$，使得

$$u(\lambda)m(\lambda)+v(\lambda)(\lambda-\lambda_0)=1$$

从而有

$$u(\boldsymbol{A})m(\boldsymbol{A})+v(\boldsymbol{A})(\boldsymbol{A}-\lambda_0\boldsymbol{E})=\boldsymbol{E}$$

注意到 $m(\boldsymbol{A})=0$，则 $v(\boldsymbol{A})(\boldsymbol{A}-\lambda_0\boldsymbol{E})=\boldsymbol{E}$. 两边取行列式可得 $|\boldsymbol{A}-\lambda_0\boldsymbol{E}|\neq 0$，与 λ_0 是 $\boldsymbol{A}$ 的特征值相矛盾. 所以，λ_0 必定是 $\boldsymbol{A}$ 的最小多项式的根.

由定理 3.4.2、定理 3.4.2 可知，若 $\boldsymbol{A}$ 的特征多项式

$$|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_s)^{m_s}$$

其中 $\sum\limits_{i=1}^{s}m_i=n$，$m_i>0$，并且当 $i\neq j$ 时，$\lambda_i\neq\lambda_j$，则

$$m(\lambda)=(\lambda-\lambda_1)^{l_1}(\lambda-\lambda_2)^{l_2}\cdots(\lambda-\lambda_s)^{l_s}$$

其中 $1\leqslant l_i\leqslant m_i$，$i=1,2,\cdots,s$.

推论 3.4.2 如果矩阵 $\boldsymbol{A}$ 的特征值互不相同，则其最小多项式就是它的特征多项式.

【例 3.4.3】 求矩阵 $\boldsymbol{A}=\begin{pmatrix}8&-3&6\\3&-2&0\\-4&2&-2\end{pmatrix}$ 的最小多项式.

解 $|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-1)^2(\lambda-2)$，因此 $\boldsymbol{A}$ 的最小多项式是

$$m_1(\lambda)=(\lambda-1)(\lambda-2) \text{ 或 } m_2(\lambda)=(\lambda-1)^2(\lambda-2)$$

∵

$$(\boldsymbol{A}-\boldsymbol{E})(\boldsymbol{A}-2\boldsymbol{E})=\begin{pmatrix}7&-3&6\\3&-3&0\\-4&2&-3\end{pmatrix}\begin{pmatrix}6&-3&6\\3&-4&0\\-4&2&-4\end{pmatrix}=\begin{pmatrix}9&3&18\\9&3&18\\-6&-2&-12\end{pmatrix}\neq 0$$

∴$\boldsymbol{A}$ 的最小多项式是 $m(\lambda)=(\lambda-1)^2(\lambda-2)$.

【例 3.4.4】 求矩阵 $\boldsymbol{A}=\begin{pmatrix}-1&2&2&0\\3&-1&1&0\\2&2&-1&0\\1&-4&3&3\end{pmatrix}$ 的最小多项式.

解 $$|A-\lambda E|=\begin{vmatrix}-1-\lambda & 2 & 2 & 0\\ 3 & -1-\lambda & 1 & 0\\ 2 & 2 & -1-\lambda & 0\\ 1 & -4 & 3 & 3-\lambda\end{vmatrix}$$

$$=(3-\lambda)\begin{vmatrix}-1-\lambda & 2 & 2\\ 3 & -1-\lambda & 1\\ 2 & 2 & -1-\lambda\end{vmatrix}$$

$$=(3-\lambda)\begin{vmatrix}3-\lambda & 2 & 2\\ 3-\lambda & -1-\lambda & 1\\ 3-\lambda & 2 & -1-\lambda\end{vmatrix}=(3-\lambda)^2\begin{vmatrix}1 & 2 & 2\\ 1 & -1-\lambda & 1\\ 1 & 2 & -1-\lambda\end{vmatrix}$$

$$=(3-\lambda)^2\begin{vmatrix}1 & 2 & 2\\ 0 & -3-\lambda & -1\\ 0 & 0 & -3-\lambda\end{vmatrix}=(\lambda-3)^2(\lambda+3)^2.$$

∵

$$(A-3E)(A+3E)=\begin{pmatrix}2 & 0 & -2 & 0\\ -4 & 0 & 4 & 0\\ 2 & 0 & -2 & 0\\ -4 & 0 & 4 & 0\end{pmatrix}\neq 0$$

$$(A-3E)^2(A+3E)=\begin{pmatrix}-12 & 0 & 12 & 0\\ 24 & 0 & -24 & 0\\ -12 & 0 & 12 & 0\\ 24 & 0 & -24 & 0\end{pmatrix}\neq 0$$

$$(A-3E)(A+3E)^2=0$$

∴A 的最小多项式 $m(\lambda)=(\lambda+3)^2(\lambda-3)$.

定理 3.4.4 n 阶矩阵 A 的最小多项式 $m(\lambda)$ 唯一存在，它等于它的特征矩阵 $\lambda E-A$ 中第 n 个不变因子 $d_n(\lambda)$，

证 唯一性：设 $m(\lambda)$、$\varphi(\lambda)$ 都是 A 的最小多项式，由定理 3.4.2

$$\varphi(\lambda)\mid m(\lambda),m(\lambda)\mid\varphi(\lambda)$$

从而 $\varphi(\lambda)$，$m(\lambda)$ 是同次多项式．又由于 $\varphi(\lambda)$，$m(\lambda)$ 的最高次幂系数都是 1，因此 $\varphi(\lambda)=m(\lambda)$.

设 λ_1，λ_2，…，λ_s 是 A 的全部相异特征值，$\lambda E-A$ 的不变因子有如下分解形式

$$\begin{cases}d_1(\lambda)=(\lambda-\lambda_1)^{e_{11}}(\lambda-\lambda_2)^{e_{12}}\cdots(\lambda-\lambda_s)^{e_{1s}}\\ d_2(\lambda)=(\lambda-\lambda_1)^{e_{21}}(\lambda-\lambda_2)^{e_{22}}\cdots(\lambda-\lambda_s)^{e_{2s}}\\ \cdots\\ d_n(\lambda)=(\lambda-\lambda_1)^{e_{n1}}(\lambda-\lambda_2)^{e_{n2}}\cdots(\lambda-\lambda_s)^{e_{ns}}\end{cases}\tag{3.4.1}$$

则有

$$e_{ij}\leqslant e_{tj},\text{其中 }1\leqslant i<t\leqslant n,j=1,2,\cdots,s;$$

又设 A 的最小多项式

$$m(\lambda)=(\lambda-\lambda_1)^{l_1}(\lambda-\lambda_2)^{l_2}\cdots(\lambda-\lambda_s)^{l_s}$$

下面证：$e_{nj}=l_j$，$j=1, 2, \cdots, s$.

由 $\lambda E-A$ 的初等因子，有 A 的 Jordan 标准形为

$$J=\mathrm{diag}(J_{11},J_{21},\cdots,J_{n1},J_{12},J_{22},\cdots,J_{n2},\cdots,J_{1s},J_{2s},\cdots,J_{ns})$$

其中

$$J_{ij}=\begin{pmatrix}\lambda_j & 1 & \cdots & 0 & 0\\ 0 & \lambda_j & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & \lambda_j & 1\\ 0 & 0 & \cdots & 0 & \lambda_j\end{pmatrix}_{e_{ij}\times e_{ij}}$$

$$(J-\lambda_kE)^m=\mathrm{diag}((J_{11}-\lambda_kE)^m,\cdots,(J_{n1}-\lambda_kE)^m,\cdots,(J_{1s}-\lambda_kE)^m,\cdots,(J_{ns}-\lambda_kE)^m)$$

当 $j=k$ 时

(1) 如果 $m\geqslant e_{nk}$，$(J_{ij}-\lambda_kE)^m=0$，$i=1, 2, \cdots, n$;

(2) 如果 $m<e_{nk}$，$(J_{nj}-\lambda_kE)^m\neq 0$.

当 $j\neq k$ 时，$(J_{ij}-\lambda_kE)^m$ 非奇异．因此 $m(\lambda)=(\lambda-\lambda_1)^{l_1}(\lambda-\lambda_2)^{l_2}\cdots(\lambda-\lambda_s)^{l_s}$ 是 A 的最小多项式充分必要条件是

$$e_{nj}=l_j, j=1,2,\cdots,s$$

即

$$m(\lambda)=d_n(\lambda)$$

定理 3.4.5 相似矩阵有相同的最小多项式，反之不然．

证 设矩阵 A 相似于矩阵 B，则存在可逆矩阵 P，使得 $A=P^{-1}BP$，如果 $f(B)=0$，那么 $f(A)=f(P^{-1}BP)=P^{-1}f(B)P=0$.

反之，不妨设 $A=\begin{pmatrix}2 & & \\ & 3 & \\ & & 3\end{pmatrix}$，$B=\begin{pmatrix}2 & & \\ & 2 & \\ & & 3\end{pmatrix}$，则 A 与 B 不相似，但是 $\det(\lambda E-A)=(\lambda-2)(\lambda-3)^2$，$\det(\lambda E-B)=(\lambda-3)(\lambda-2)^2$，所以 A 与 B 有相同的最小多项式 $m(A)=m(B)=(\lambda-2)(\lambda-3)$.

【例 3.4.5】 求矩阵 $A=\begin{pmatrix}3 & 0 & 8\\ 3 & -1 & 6\\ -2 & 0 & -5\end{pmatrix}$ 的最小多项式．

解 由例 3.3.2 知，A 的初等因子是 $\lambda+1$，$(\lambda+1)^2$，所以 A 的最小多项式为 $m(\lambda)=(\lambda+1)^2$.

由不变因子的构成，对应于求 A 的 Jordan 标准形的波尔曼算法，我们可得求最小多项式的另一种方法．其具体步骤如下：

(1) 求 A 的所有相异特征值，不妨设 A 的所有相异特征值为 $\lambda_1, \cdots, \lambda_s$;

(2) 对每个相异的特征根 λ_k 和每个 j，$(j=1, \cdots, n)$，求 $(\lambda_kE-A)^j$ 的秩

$$r_j(\lambda_k)=\mathrm{rank}(\lambda_kE-A)^j$$

在计算秩时，若发现对某个 j^*，有 $r_{j^*}(\lambda_k)=r_{j^*+1}(\lambda_k)$，则停止，并令 $a_k=j^*(\lambda_k)$;

(3) 写出 A 的最小多项式

$$\varphi(\lambda)=(\lambda-\lambda_1)^{a_1}\cdots(\lambda-\lambda_s)^{a_s}$$

【例 3.4.6】 设$A=\begin{pmatrix}-1&-2&6\\-1&0&3\\-1&-1&4\end{pmatrix}$，求 A 的最小多项式．

解 $$|\lambda E-A|=\begin{vmatrix}\lambda+1&2&-6\\1&\lambda&-3\\1&1&\lambda-4\end{vmatrix}=(\lambda-1)^3$$

$$E-A=\begin{pmatrix}2&2&-6\\1&1&-3\\1&1&-3\end{pmatrix}=\begin{pmatrix}2\\1\\1\end{pmatrix}(1\quad 1\quad -3),(E-A)^2=\mathbf{0}$$

$$r_1(1)=1,r_2(1)=0,r_3(1)=0.r_4(1)=0$$

由于$r_1(1)\neq r_2(1)=r_3(1)$，因此 A 的最小多项式是 $(\lambda-1)^2$．

若 A 的最小多项式是 m 次多项式，将 E，A，A^2，…，A^m 看成向量，则有 E，A，A^2，…，A^m 线性相关，而 E，A，A^2，…，$A^k(k<m)$ 线性无关．由此，我们可得求最小多项式的第三种方法：

(1) 试解 $A=kE$，若有解，则 $\varphi(\lambda)=\lambda-k$，否则继续下一步；

(2) 对 $i>1$，试解：$A^i=k_0I+k_1A+\cdots+k_{i-1}A^{i-1}$.

1) 若有解，则 $\varphi(\lambda)=\lambda^i-k_0-k_1\lambda-\cdots-k_{i-1}\lambda^{i-1}$；

2) 否则继续下一个 i.

【例 3.4.7】 设$A=\begin{pmatrix}-1&-2&6\\-1&0&3\\-1&-1&4\end{pmatrix}$，求 A 的最小多项式．

解 显然 $A=kE$ 无解．

$$A^2=\begin{pmatrix}-3&-4&12\\-2&-1&6\\-2&-2&7\end{pmatrix}$$

令

$$\begin{pmatrix}-3&-4&12\\-2&-1&6\\-2&-2&7\end{pmatrix}=k_1\begin{pmatrix}-1&-2&6\\-1&0&3\\-1&-1&4\end{pmatrix}+k_2\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$$

由于上方程有解 $k_1=2$，$k_2=-1$，故 A 的最小多项式是$\lambda^2-2\lambda+1$.

【例 3.4.8】 设$A=\begin{pmatrix}4&3&-4\\-1&0&2\\1&1&0\end{pmatrix}$，求 A 的最小多项式和 Jordan 标准型．

解 $|\lambda E-A|=(\lambda-1)^2(\lambda-2)$

$$(A-E)(A-2E)=\begin{pmatrix}3&3&-4\\-1&-1&2\\1&1&-1\end{pmatrix}\begin{pmatrix}2&3&-4\\-1&-2&2\\1&1&-2\end{pmatrix}=\begin{pmatrix}-1&-1&2\\1&1&-2\\0&0&0\end{pmatrix}\neq\mathbf{0}$$

因此 A 的最小多项式是$m(\lambda)=(\lambda-1)^2(\lambda-2)$，Jordan 标准型 $J=\begin{pmatrix}1&1&0\\0&1&0\\0&0&2\end{pmatrix}$.

定理 3.4.6 A 可以对角化的充分必要条件是 A 的最小多项式无重根．

证 必要性：设 A 的相异特征值是 λ_1，λ_2，…，λ_s．因为 A 可以对角化，所以 A 的 Jordan标准形为

$$J=\begin{pmatrix}\lambda_1 E_{k_1} & & & \\ & \lambda_2 E_{k_2} & & \\ & & \ddots & \\ & & & \lambda_s E_{k_s}\end{pmatrix}$$

因此 $\lambda E-A$ 的初等因子全是关于 λ 的一次因式，故 A 的最小多项式无重根．

充分性：由于 A 的最小多项式无重根，因此 $\lambda E-A$ 的初等因子全是关于 λ 的一次因式．由此有 A 的 Jordan 块都是一阶块，即 A 可以对角化．

推论 3.4.3 若 A 的零化多项式无重根，则 A 可以对角化．

证 由于 A 的零化多项式能被 A 的最小多项式整除，因此若 A 的零化多项式无重根，则它的最小多项式也无重根，由定理 3.4.6 知 A 可以对角化．

【例 3.4.9】 设 n 阶矩阵 A 满足 $A^2-2A-3E=0$，且 $\operatorname{rank}(A-3E)=r$，求 A 的 Jordan 标准型．

解 因为矩阵 A 满足 $A^2-2A-3E=0$，所以 $\lambda^2-2\lambda-3$ 是 A 的零化多项式，它无重根，因此 A 可以对角化，即 A 的 Jordan 标准形有如下形式

$$J=\begin{pmatrix}3E_s & \\ & -E_{n-s}\end{pmatrix}$$

由于

$$\operatorname{rank}(J-3E)=\operatorname{rank}(A-3E)=r$$

故 $s=n-r$. 因此 A 的 Jordan 标准型为

$$J=\begin{pmatrix}3E_{n-r} & \\ & -E_r\end{pmatrix}$$

【例 3.4.10】 求 $A=\begin{pmatrix}1 & 1 & \cdots & 1\\ 1 & 1 & \cdots & 1\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & \cdots & 1\end{pmatrix}_{n\times n}$ 的最小多项式．

解 由于 A 是实对称阵，它能对角化．因而它的最小多项式无重根．

由 Cayley-Hamilton 定理知 A 的特征多项式 $|\lambda E-A|=\lambda^{n-1}(\lambda-n)$ 是 A 的零化多项式，它能被 A 的最小多项式整除．注意到 A 的特征多项式与 A 的最小多项式有相同的零点，因此 A 的最小多项式是 $\lambda(\lambda-n)$．

习 题 3

3.1 下列的 λ-矩阵中哪些是满秩的，哪些是可逆的？如果是可逆时求其逆方阵．

(1) $\begin{pmatrix} \lambda & 2\lambda+1 & 1 \\ 1 & \lambda+1 & \lambda^2+1 \\ \lambda-1 & \lambda & -\lambda^2 \end{pmatrix}$；　(2) $\begin{pmatrix} 1 & 0 & 1 \\ 0 & \lambda-1 & \lambda \\ \lambda & 1 & \lambda^2 \end{pmatrix}$；　(3) $\begin{pmatrix} 1 & \lambda & 0 \\ 2 & \lambda & 1 \\ \lambda^2+1 & 2 & \lambda^2+1 \end{pmatrix}$.

3.2 化下列矩阵为 Smith 标准形，并求出它们的不变因子和初等因子.

(1) $\begin{pmatrix} 1-\lambda & \lambda^2 & \lambda \\ \lambda & \lambda & -\lambda \\ 1+\lambda^2 & \lambda^2 & -\lambda^2 \end{pmatrix}$；　(2) $\begin{pmatrix} 3\lambda^2+2\lambda-3 & 2\lambda-1 & \lambda^2+2\lambda-3 \\ 4\lambda^2+3\lambda-5 & 3\lambda-2 & \lambda^2+3\lambda-4 \\ \lambda^2+\lambda-4 & \lambda-2 & \lambda-1 \end{pmatrix}$；

(3) $\begin{pmatrix} 2\lambda & 3 & 0 & 1 & \lambda \\ 4\lambda & 3\lambda+6 & 0 & \lambda+2 & 2\lambda \\ 0 & 6\lambda & \lambda & 2\lambda & 0 \\ \lambda-1 & 0 & \lambda-1 & 0 & 0 \\ 3\lambda-3 & 1-\lambda & 2\lambda-2 & 0 & 0 \end{pmatrix}$；　(4) $\begin{pmatrix} \lambda-3 & -1 & 0 & 0 \\ 4 & \lambda+1 & 0 & 0 \\ -6 & -1 & \lambda-2 & -1 \\ 14 & 5 & 1 & \lambda \end{pmatrix}$；

(5) $\begin{pmatrix} 0 & 0 & 2 & \lambda+3 \\ 0 & -1 & \lambda-3 & 0 \\ 1 & \lambda+3 & \lambda+3 & 3 \\ \lambda-3 & \lambda-5 & 0 & \lambda^3+2\lambda-6 \end{pmatrix}$；　(6) $\begin{pmatrix} 1-\lambda & \lambda^2 & \lambda \\ \lambda & \lambda & -\lambda \\ 1+\lambda^2 & \lambda^2 & -\lambda^2 \end{pmatrix}$；

(7) $\begin{pmatrix} \lambda^2-4 & 0 & 0 & 0 \\ 0 & \lambda^2-5\lambda+6 & 0 & 0 \\ 0 & 0 & 0 & \lambda+2 \\ 0 & 0 & \lambda-2 & 0 \end{pmatrix}$；　(8) $\begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 1 & \lambda & 0 \\ 0 & 0 & 1 & \lambda \end{pmatrix}$.

3.3 证明：任何方阵 $\boldsymbol{A}$ 与其转置矩阵 $\boldsymbol{A}^{\mathrm{T}}$ 必相似.

3.4 设 $\boldsymbol{A}(\lambda)$ 是一个 5×6 阶的 λ-矩阵，秩为 4，初等因子为

$$\lambda,\lambda,\lambda^2,\lambda-1,\lambda-1,(\lambda+1)^2$$

试求 $\boldsymbol{A}(\lambda)$ 的标准形.

3.5 求下列矩阵的 Jordan 标准形：

(1) $\begin{pmatrix} 3 & 1 & -3 \\ -7 & -2 & 9 \\ -2 & -1 & 4 \end{pmatrix}$；　(2) $\begin{pmatrix} 3 & 3 & 2 \\ 1 & 1 & -2 \\ -3 & -1 & 0 \end{pmatrix}$；

(3) $\begin{pmatrix} 3 & -3 & 2 \\ -1 & 5 & -2 \\ -1 & 3 & 0 \end{pmatrix}$；　(4) $\begin{pmatrix} 7 & 4 & -4 \\ 4 & -8 & -1 \\ -4 & -1 & -8 \end{pmatrix}$；

(5) $\begin{pmatrix} 4 & -2 & 2 \\ -5 & 7 & -5 \\ -6 & 7 & -4 \end{pmatrix}$；　(6) $\begin{pmatrix} 3 & 7 & -3 \\ -2 & -5 & 2 \\ -4 & -10 & 3 \end{pmatrix}$.

3.6 求下列矩阵的最小多项式：

(1) $\begin{pmatrix} 2 & 6 & -15 \\ 1 & 1 & -5 \\ 1 & 2 & -6 \end{pmatrix}$；　(2) $\begin{pmatrix} 4 & 6 & -15 \\ 1 & 3 & -5 \\ 1 & 2 & -4 \end{pmatrix}$；

(3) $\begin{bmatrix} 5 & -3 & 2 \\ 6 & -4 & 4 \\ 4 & -4 & 5 \end{bmatrix}$； (4) $\begin{bmatrix} 4 & -5 & 2 \\ 5 & -7 & 3 \\ 6 & -9 & 4 \end{bmatrix}$；

(5) $\begin{bmatrix} 1 & -3 & 3 \\ -2 & -6 & 13 \\ -1 & -4 & 8 \end{bmatrix}$； (6) $\begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$.

3.7 求下列矩阵 $\boldsymbol{A}$ 的 Jordan 标准形 $\boldsymbol{J}$，并求相应的相似变换矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}$.

(1) $\boldsymbol{A}=\begin{bmatrix} 1 & -3 & 3 \\ -2 & -6 & 13 \\ -1 & -4 & 8 \end{bmatrix}$； (2) $\begin{bmatrix} 4 & 2 & -5 \\ 6 & 1 & -9 \\ 5 & 3 & -7 \end{bmatrix}$.

3.8 设线性变换 T 在三维线件空间 V 的基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\boldsymbol{\alpha}_3$ 下的矩阵为

$$\boldsymbol{A}=\begin{bmatrix} 5 & -3 & 2 \\ 6 & -4 & 4 \\ 4 & -4 & 5 \end{bmatrix}$$

求 V 的另一组基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3$，使 T 在该基下的矩阵为 Jordan 标准形.

3.9 设 $\boldsymbol{A}=\begin{bmatrix} 1 & -3 & 4 \\ 4 & -7 & 8 \\ 6 & -7 & 7 \end{bmatrix}$，求：

(1) 矩阵 $\boldsymbol{A}$ 的 Jordan 标准形 $\boldsymbol{J}$，并求相应的相似变换矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}$；

(2) 矩阵 $\boldsymbol{A}$ 的最小多项式；

(3) 计算 $\boldsymbol{A}^{10}-2\boldsymbol{A}^7+3\boldsymbol{A}^5-5\boldsymbol{A}+6\boldsymbol{E}$.

3.10 证明：矩阵 $\boldsymbol{A}$ 可逆的充分必要条件是 $\lambda=0$ 不是 $\boldsymbol{A}$ 的最小多项式的零点.

3.11 证明：矩阵 $\boldsymbol{A}$ 是幂零阵，即存在自然数 k，使得 $\boldsymbol{A}^k=\boldsymbol{0}$，当且仅当它的所有特征值等于零.

3.12 求幂等矩阵 $\boldsymbol{A}$ 的 Jordan 标准形，以及它的最小多项式.

3.13 设 $\boldsymbol{A}^2=\boldsymbol{0}$，$\mathrm{rank}(\boldsymbol{A})=r<n$，求 $\boldsymbol{A}$ 的 Jordan 标准形.

3.14 设 5 阶实矩阵 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^3(\boldsymbol{A}-2\boldsymbol{E})=\boldsymbol{0}$，$\mathrm{rank}(\boldsymbol{A})=3$，求 $\boldsymbol{A}$ 的 Jordan 标准形.

3.15 设 $\boldsymbol{A}^2=\boldsymbol{E}$，$\mathrm{rank}(\boldsymbol{A}+\boldsymbol{E})=r$，求 $|\boldsymbol{A}-2\boldsymbol{E}|$.

3.16 设 $\boldsymbol{A}$ 是一个 6 阶方阵，它的特征多项式是 $f(\lambda)=(\lambda-2)^4(\lambda-3)^2$，求：

(1) 如果 $\boldsymbol{A}$ 的最小多项式分别为 $m(\lambda)=(\lambda-2)^2(\lambda-3)^2$，试写出矩阵可能的 Jordan 标准形；

(2) 如果 $\boldsymbol{A}$ 的最小多项式分别为 $m(\lambda)=(\lambda-2)(\lambda-3)^2$，试写出矩阵可能的 Jordan 标准形.

3.17 已知 $\boldsymbol{A}=\begin{bmatrix} 2 & 0 & 0 \\ a & 2 & 0 \\ b & c & -1 \end{bmatrix}$，求

(1) $\boldsymbol{A}$ 的一切可能的 Jordan 标准形；

(2) $\boldsymbol{A}$ 对角化的一个充要条件.

3.18 设 $\boldsymbol{A}$ 是 n 阶方阵，且 0 是 $\boldsymbol{A}$ 的 k 重特征根. 证明：当且仅当 $\mathrm{rank}(\boldsymbol{A})=n-k$ 时，

$\text{rank}(\boldsymbol{A}^2) = \text{rank}(\boldsymbol{A})$.

3.19 设 $\boldsymbol{J}$ 是特征根为 λ 的 n 阶 Jordan 块，求：

(1) 与 $\boldsymbol{J}$ 可换的矩阵；

(2) 证明：与 $\boldsymbol{J}$ 可换的矩阵必为 $\boldsymbol{J}$ 的多项式.

3.20 设 T 是复数域上的 n 维空间 $\mathbf{C}^n$ 的一个线性变换，而 T 在基 $\boldsymbol{e}_1$，$\boldsymbol{e}_2$，…，$\boldsymbol{e}_n$ 下的矩阵是一个约当块. 证明：

(1) 包含 $\boldsymbol{e}_1$ 的不变子空间就是 $\mathbf{C}^n$ 自身；

(2) 任一非零不变子空间都包含 $\boldsymbol{e}_n$；

(3) $\mathbf{C}^n$ 不能分解成两个非平凡的不变子空间的直和.

第四章 向量与矩阵的范数

向量（或矩阵）的范数是向量（或矩阵）的数字特征. 范数在讨论向量（或矩阵）序列和矩阵级数的收敛时起着根本的作用.

§4.1 向量范数

我们知道在三维欧氏空间中，向量的长度可以按如下方法定义：假设 $\boldsymbol{\alpha}\in\mathbf{R}^3$，并且 $\boldsymbol{\alpha}=(x_1,x_2,x_3)$，则向量 $\boldsymbol{\alpha}$ 的长度为 $\sqrt{x_1^2+x_2^2+x_3^2}$. 但是对于 n 维欧氏空间 $\mathbf{R}^n$，n 维复空间 $\mathbf{C}^n$ 或一般的线性空间，应该如何定义向量的长度呢？为此引入范数的概念.

定义 4.1.1 设 $\boldsymbol{V}$ 是数域 P 上的线性空间，如果对于任意的向量 $\boldsymbol{x}\in\boldsymbol{V}$，都对应着一个实数 $N(\boldsymbol{x})=\|\boldsymbol{x}\|$，并且满足下列条件：

(1) $\|\boldsymbol{x}\|\geqslant 0$，$\|\boldsymbol{x}\|=0$ 当且仅当 $\boldsymbol{x}=\mathbf{0}$（正定条件）；

(2) $\|k\boldsymbol{x}\|=|k|\,\|\boldsymbol{x}\|$，$\forall k\in P$，$\boldsymbol{x}\in\boldsymbol{V}$（齐次条件）；

(3) $\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|$（三角不等式）.

则称实数 $N(\boldsymbol{x})=\|\boldsymbol{x}\|$ 是 $\boldsymbol{V}$ 上的一个向量范数（或称为模）.

实际上，$\mathbf{R}^3$ 中向量 $\boldsymbol{\alpha}$ 的长度 $\sqrt{x_1^2+x_2^2+x_3^2}$ 就是 $\mathbf{R}^3$ 中的一个向量范数，它满足定义 4.1.1 所列的三个条件. 所以范数是比长度更广义的一个概念.

由向量范数定义不难验证向量范数具有下列性质：

(1) 当 $\|\boldsymbol{x}\|\neq 0$ 时，$\left\|\dfrac{1}{\|\boldsymbol{x}\|}\boldsymbol{x}\right\|=1$；

(2) $\forall\boldsymbol{x}\in\boldsymbol{V}$，$\|-\boldsymbol{x}\|=\|\boldsymbol{x}\|$；

(3) $\forall\boldsymbol{x}$，$\boldsymbol{y}\in\boldsymbol{V}$，$\|\boldsymbol{x}\|-\|\boldsymbol{y}\|\leqslant\|\boldsymbol{x}-\boldsymbol{y}\|$；

(4) $\forall\boldsymbol{x}$，$\boldsymbol{y}\in\boldsymbol{V}$，$|\,\|\boldsymbol{x}\|-\|\boldsymbol{y}\|\,|\leqslant\|\boldsymbol{x}+\boldsymbol{y}\|$.

【例 4.1.1】 设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\mathbf{C}^n$，规定 $\|\boldsymbol{x}\|=\max\limits_i|x_i|$，证明 $\|\boldsymbol{x}\|$ 是 $\mathbf{C}^n$ 中的一个向量范数. 这个范数称为 $\boldsymbol{x}$ 的 ∞-范数. 记作 $\|\boldsymbol{x}\|_\infty=\max\limits_i|x_i|$.

证 (1) 当 $\boldsymbol{x}\neq 0$，有 $\|\boldsymbol{x}\|=\max\limits_i|x_i|>0$，当 $\boldsymbol{x}=\mathbf{0}$ 时 $\|\boldsymbol{x}\|=0$；

(2) $\forall k\in\mathbf{C}$, $\|k\boldsymbol{x}\|_\infty=\max\limits_i|k\boldsymbol{x}_i|=|k|\max\limits_i|\boldsymbol{x}_i|=|k|\,\|\boldsymbol{x}\|_\infty$；

(3) $\forall\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，$\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}\in\mathbf{C}^n$，有

$$\|\boldsymbol{x}+\boldsymbol{y}\|_\infty=\max_i|x_i+y_i|\leqslant\max_i|x_i|+\max_i|y_i|=\|\boldsymbol{x}\|_\infty+\|\boldsymbol{y}\|_\infty$$

所以 $\|\boldsymbol{x}\|_\infty=\max\limits_i|x_i|$ 是 $\mathbf{C}^n$ 中的一个向量范数.

【例 4.1.2】 设 $\boldsymbol{x}=(x_1,x_2,\cdots x_n)^{\mathrm{T}}\in\mathbf{C}^n$，规定 $\|\boldsymbol{x}\|=\sum\limits_{i=1}^{n}|x_i|$，证明 $\|\boldsymbol{x}\|$ 是 $\mathbf{C}^n$ 中的一个向量范数. 这个范数称为 x 的 1-范数，记作 $\|\boldsymbol{x}\|_1=\sum\limits_{i=1}^{n}|x_i|$.

证 (1) 当 $\boldsymbol{x}\neq\boldsymbol{0}$ 时，x_1，x_2，…，x_n 必不全为零，所以$\|\boldsymbol{x}\|_1=|x_1|+|x_2|+\cdots+|x_n|>0$，当 $\boldsymbol{x}=\boldsymbol{0}$ 时，必有 $x_1=x_2=\cdots=x_n=0$，所以$\|\boldsymbol{x}\|_1=0$；

(2) $\forall k\in\mathbf{C},\|k\boldsymbol{x}\|_1=\sum\limits_{i=1}^{n}|kx_i|=|k|\sum\limits_{i=1}^{n}|x_i|=|k|\,\|\boldsymbol{x}\|_1$；

(3) $\forall\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}},\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}\in\mathbf{C}^n$，有

$$\|\boldsymbol{x}+\boldsymbol{y}\|_1=\sum_{i=1}^{n}|x_i+y_i|\leqslant\sum_{i-1}^{n}|x_i|+\sum_{i=1}^{n}|y_i|=\|\boldsymbol{x}\|_1+\|\boldsymbol{y}\|_1$$

所以 $\|\boldsymbol{x}\|_1=\sum\limits_{i=1}^{n}|x_i|$ 是 $\mathbf{C}^n$ 中的一个向量范数.

【例 4.1.3】 设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\mathbf{C}^n$，规定 $\|\boldsymbol{x}\|=\sqrt{\sum\limits_{i=1}^{n}|x_i|^2}$，证明 $\|\boldsymbol{x}\|$是 $\mathbf{C}^n$ 中的一个向量范数. 这个范数称为 $\boldsymbol{x}$ 的 2-范数. 记作 $\|\boldsymbol{x}\|_2=\sqrt{\sum\limits_{i=1}^{n}|x_i|^2}$.

证 (1) 当 $\boldsymbol{x}\neq\boldsymbol{0}$ 时，x_1，x_2，…，x_n 必不全为零，所以 $\|\boldsymbol{x}\|_2=\sqrt{\sum\limits_{i=1}^{n}|x_i|^2}>0$，当 $\boldsymbol{x}=\boldsymbol{0}$ 时，必有 $x_1=x_2=\cdots=x_n=0$，所以 $\|\boldsymbol{x}\|_2=\sqrt{\sum\limits_{i=1}^{n}|x_i|^2}=0$；

(2) $\forall k\in\mathbf{C},\|k\boldsymbol{x}\|_2=\sqrt{\sum\limits_{i=1}^{n}|kx_i|^2}=|k|\sqrt{\sum\limits_{i=1}^{n}|x_i|^2}=|k|\,\|\boldsymbol{x}\|_2$；

(3) $\forall\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}},\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}\in\mathbf{C}^n$，有

$$\|\boldsymbol{x}+\boldsymbol{y}\|_2=\sqrt{\sum_{i=1}^{n}|x_i+y_i|^2}$$

由 $\mathbf{C}^n$ 中向量的内积可知

$$\|\boldsymbol{x}+\boldsymbol{y}\|_2^2=(\boldsymbol{x}+\boldsymbol{y},\boldsymbol{x}+\boldsymbol{y})=(\boldsymbol{x},\boldsymbol{x})+2\mathrm{Re}(\boldsymbol{x},\boldsymbol{y})+(\boldsymbol{y},\boldsymbol{y})$$

又因为 $\mathrm{Re}(\boldsymbol{x},\boldsymbol{y})\leqslant|(\boldsymbol{x},\boldsymbol{y})|\leqslant\sqrt{(\boldsymbol{x},\boldsymbol{x})(\boldsymbol{y},\boldsymbol{y})}=\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2$，所以 $\|\boldsymbol{x}+\boldsymbol{y}\|_2^2\leqslant\|\boldsymbol{x}\|_2^2+2\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2+\|\boldsymbol{y}\|_2^2=(\|\boldsymbol{x}\|_2+\|\boldsymbol{y}\|_2)^2$，则$\|\boldsymbol{x}+\boldsymbol{y}\|_2\leqslant\|\boldsymbol{x}\|_2+\|\boldsymbol{y}\|_2$. 所以 $\|\boldsymbol{x}\|_2=\sqrt{\sum\limits_{i=1}^{n}|x_i|^2}$ 是 $\mathbf{C}^n$ 中的一个向量范数. 2-范数也叫欧氏范数.

【例 4.1.4】 设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\mathbf{C}^n$，规定 $\|\boldsymbol{x}\|=\left(\sum\limits_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}$，$(1\leqslant p<+\infty)$，证明$\|\boldsymbol{x}\|$是 $\mathbf{C}^n$ 中的一个向量范数. 这个范数称为 $\boldsymbol{x}$ 的 p-范数. 记作 $\|\boldsymbol{x}\|_p=\left(\sum\limits_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}$.

证 (1) 当 $\boldsymbol{x}\neq\boldsymbol{0}$ 时，x_1，x_2，…，x_n 必不全为零，所以 $\|\boldsymbol{x}\|_p=\left(\sum\limits_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}>0$，当 $\boldsymbol{x}=\boldsymbol{0}$ 时，必有 $x_1=x_2=\cdots=x_n=0$，所以 $\|\boldsymbol{x}\|_p=\left(\sum\limits_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}=0$；

(2) $\forall k\in\mathbf{C}$，$\|k\boldsymbol{x}\|_p=\left(\sum\limits_{i=1}^{n}|kx_i|^p\right)^{\frac{1}{p}}=|k|\left(\sum\limits_{i=1}^{n}|kx_i|^p\right)^{\frac{1}{p}}=|k|\,\|\boldsymbol{x}\|_p$；

(3) $\forall\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}},\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}\in\mathbf{C}^n$，有

$$\|x+y\|_p=\left(\sum_{i=1}^n|x_i+y_i|^p\right)^{\frac{1}{p}}\leqslant\left(\sum_{i=1}^n|x_i|^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^n|y_i|^p\right)^{\frac{1}{p}}=\|x\|_p+\|y\|_p$$

所以 $\|x\|_p=\left(\sum\limits_{i=1}^n|x_i|^p\right)^{\frac{1}{p}}$ 是 $\mathbf{C}^n$ 中的一个向量范数.

显然，$p=1$，2 时，p-范数就是 1-范数和 2-范数，并且 $\lim\limits_{p\to+\infty}\|x\|_p=\|x\|_\infty$.

由以上例题可见，在一个线性空间中，可以定义多种向量范数. 那么这多种向量范数之间有什么关系呢?

定义 4.1.2　若 $\|x\|_\alpha$ 和 $\|x\|_\beta$ 是维线性空间 $\mathbf{V}$ 中的任意两个向量范数，如果存在 $M>0$，$m>0$，使得对 $\forall x\in\mathbf{V}$，都有 $m\|x\|_\beta\leqslant\|x\|_\alpha\leqslant M\|x\|_\beta$，则称范数 $\|x\|_\alpha$ 与 $\|x\|_\beta$ 等价.

定理 4.1.1　有限维线性空间中任意两个向量范数都等价.

证　取定 n 维线性空间 $\mathbf{V}^n$ 的一个基 e_1，e_2，…，e_n，则对 $\forall x\in\mathbf{V}^n$，$x=x_1e_1+x_2e_2+\cdots+x_ne_n$.

首先，证明 $\mathbf{V}^n$ 中任意一个向量范数 $\|x\|$ 都是坐标 x_1，x_2，…，x_n 的连续函数. 令 $\varphi(x_1,x_2,\cdots,x_n)=\|x\|$，则

$$\begin{aligned}|\varphi(x_1,x_2,\cdots,x_n)-\varphi(y_1,y_2,\cdots,y_n)|&=|\|x\|-\|y\||\leqslant\|x-y\|\\&=\left\|\sum_{i=1}^n(x_i-y_i)e_i\right\|\leqslant\sum_{i=1}^n(|x_i-y_i|\|e_i\|)\\&\leqslant\sqrt{\sum_{i=1}^n\|e_i\|^2}\sqrt{\sum_{i=1}^n|x_i-y_i|^2}=k\sqrt{\sum_{i=1}^n|x_i-y_i|^2}\end{aligned}$$

当 $x\to y$ 时，$|\varphi(x_1,x_2,\cdots,x_n)-\varphi(y_1,y_2,\cdots,y_n)|\to0$，所以，$\varphi(x_1,x_2,\cdots,x_n)=\|x\|$ 是坐标 x_1，x_2，…，x_n 的连续函数.

其次，设 $\|x\|_\alpha$ 和 $\|x\|_\beta$ 是有限维线性空间 $\mathbf{V}^n$ 中的任意两个向量范数，欲证明存在 $M>0$，$m>0$，使得对 $\forall x\in\mathbf{V}^n$，都有 $m\|x\|_\beta\leqslant\|x\|_\alpha\leqslant M\|x\|_\beta$. 当 $x=\mathbf{0}$ 时，上式显然成立. 当 $x\neq\mathbf{0}$ 时，则有 $\|x\|_\beta\neq\mathbf{0}$，因为 $\|x\|_\alpha$ 和 $\|x\|_\beta$ 都是坐标 x_1，x_2，…，x_n 的连续函数，所以 $f(x_1,x_2,\cdots,x_n)=\dfrac{\|x\|_\alpha}{\|y\|_\beta}$ 也是坐标 x_1，x_2，…，x_n 的连续函数. 考虑 n 维线性空间 $\mathbf{V}^n$ 中的单位球面 $S=\{(\xi_1,\xi_2,\cdots,\xi_n):|\xi_1|^2+|\xi_2|^2+\cdots+|\xi_n|^2=1\}$，因为 S 是有限闭集，并且 S 上的点都不为零，所以 $f(x_1,x_2,\cdots,x_n)$ 是 S 上的连续函数. 因此，$f(x_1,x_2,\cdots,x_n)$ 在 S 上取得最大值 M 与最小值 m，亦即存在 x_0，$y_0\in S$，使得

$$M=\max_{x\in S}\frac{\|x\|_\alpha}{\|x\|_\beta}=\frac{\|x_0\|_\alpha}{\|x_0\|_\beta}>0,\quad m=\min_{x\in S}\frac{\|x\|_\alpha}{\|x\|_\beta}=\frac{\|y_0\|_\alpha}{\|y_0\|_\beta}>0$$

那么存在着 $\xi=\dfrac{x}{\sqrt{\sum\limits_{i=1}^n|x_i|^2}}\in S$，使得

$$m\leqslant\frac{\|\xi\|_\alpha}{\|\xi\|_\beta}=\frac{\|x\|_\alpha}{\|x\|_\beta}\leqslant M$$

所以

$$m\|x\|_\beta\leqslant\|x\|_\alpha\leqslant M\|x\|_\beta$$

下面我们来讨论如何由已知范数构造新的范数.

【例 4.1.5】　设 $\|y\|_\alpha$ 是 $\mathbf{C}^m$ 中的一个向量范数，给定矩阵 $A\in\mathbf{C}^{m\times n}$，它的 n 个列向量

线性无关. 对于 $\mathbf{C}^n$ 中的任意向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，规定 $\|\boldsymbol{x}\|_\beta=\|\boldsymbol{Ax}\|_\alpha$，则 $\|\boldsymbol{x}\|_\beta$ 也是 $\mathbf{C}^n$ 中的一个向量范数.

证 (1) 设 $\boldsymbol{A}=(\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_n)$，由假设 $\boldsymbol{A}_1$，$\boldsymbol{A}_2$，…，$\boldsymbol{A}_n$ 是 n 个线性无关的列向量可知，当 $\boldsymbol{x}\neq\boldsymbol{0}$ 时

$$\boldsymbol{Ax}=(\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\boldsymbol{A}_1x_1+\boldsymbol{A}_2x_2+\cdots+\boldsymbol{A}_nx_n\neq\boldsymbol{0}$$

又因为 $\|\boldsymbol{y}\|_\alpha$ 是 $\mathbf{C}^m$ 中的一个向量范数，所以 $\|\boldsymbol{Ax}\|_\alpha>0$，即 $\|\boldsymbol{x}\|_\beta>0$；当 $\boldsymbol{x}=\boldsymbol{0}$ 时，$\boldsymbol{Ax}=\boldsymbol{0}$，所以 $\|\boldsymbol{x}\|_\beta=\|\boldsymbol{Ax}\|_\alpha=0$.

(2) $\forall k\in\mathbf{C}$，$\|k\boldsymbol{x}\|_\beta=\|\boldsymbol{A}(k\boldsymbol{x})\|_\alpha=\|k\boldsymbol{Ax}\|_\alpha=|k|\,\|\boldsymbol{Ax}\|_\alpha=|k|\,\|\boldsymbol{x}\|_\beta$.

(3) $\forall\,\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}},\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}\in\mathbf{C}^n$，有

$$\|\boldsymbol{x}+\boldsymbol{y}\|_\beta=\|\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{y})\|_\alpha=\|\boldsymbol{Ax}+\boldsymbol{Ay}\|_\alpha\leqslant\|\boldsymbol{Ax}\|_\alpha+\|\boldsymbol{Ay}\|_\alpha=\|\boldsymbol{x}\|_\beta+\|\boldsymbol{y}\|_\beta$$

所以 $\|\boldsymbol{x}\|_\beta$ 是 $\mathbf{C}^n$ 中的一个向量范数.

由此可知，当给定一个 $m\times n$ 矩阵 $\boldsymbol{A}$ 时，可以由 $\mathbf{C}^m$ 中的向量范数定义 $\mathbf{C}^n$ 中的一个向量范数. 当然矩阵 $\boldsymbol{A}$ 的 n 个列向量要线性无关.

定义 4.1.3 若 $\{\boldsymbol{x}^{(k)}\}(k=1,2,\cdots)$ 是线性空间 $\boldsymbol{V}$ 中的向量序列，如果存在 $\boldsymbol{x}\in\boldsymbol{V}$，使得 $\lim\limits_{k\to+\infty}\|\boldsymbol{x}^{(k)}-\boldsymbol{x}\|_\alpha=0$，则称序列 $\{\boldsymbol{x}^{(k)}\}(k=1,2,\cdots)$ 按 α-范数收敛于 $\boldsymbol{x}$.

由于在线性空间 $\boldsymbol{V}$ 中可以定义多种范数，因此 $\boldsymbol{V}$ 中的向量序列就有多种收敛. 那么，这些按不同范数的收敛之间有什么关系呢？它与向量序列按坐标收敛的关系又如何？我们有如下定理.

定理 4.1.2 (1) 在有限维线性空间中，若向量序列 $\{\boldsymbol{x}^{(k)}\}$ 按某种范数收敛于 $\boldsymbol{x}$，则 $\{\boldsymbol{x}^{(k)}\}$ 按任何范数都收敛于 $\boldsymbol{x}$；

(2) 在有限维线性空间中，按范数收敛于 $\boldsymbol{x}$ 与按坐标收敛于 $\boldsymbol{x}$ 等价.

证 略.

§4.2 矩阵的范数

在讨论了向量范数之后，我们继续讨论矩阵的范数. 矩阵空间 $\mathbf{C}^{m\times n}$ 是一个 $m\times n$ 维的线性空间，对于 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$，可视 $\boldsymbol{A}$ 为 $\mathbf{C}^{m\times n}$ 中的一个向量，由向量范数的定义可以类似的定义矩阵的范数.

定义 4.2.1 若对于任意的矩阵 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$，都对应着一个实数 $N(\boldsymbol{A})=\|\boldsymbol{A}\|$，并且满足以下条件：

(1) 当 $\boldsymbol{A}\neq\boldsymbol{0}$ 时，$\|\boldsymbol{A}\|>0$，当且仅当 $\boldsymbol{A}=\boldsymbol{0}$ 时，$\|\boldsymbol{A}\|=0$； (正定条件)

(2) $\|k\boldsymbol{A}\|=|k|\,\|\boldsymbol{A}\|$ $(\forall k\in\mathbf{C},\ \boldsymbol{A}\in\mathbf{C}^{m\times n})$； (齐次条件)

(3) $\|\boldsymbol{A}+\boldsymbol{B}\|\leqslant\|\boldsymbol{A}\|+\|\boldsymbol{B}\|$ $(\forall\boldsymbol{A},\boldsymbol{B}\in\mathbf{C}^{m\times n})$； (三角不等式)

则实值函数 $N(\boldsymbol{A})=\|\boldsymbol{A}\|$ 被称为 $\boldsymbol{A}$ 的广义矩阵范数.

如果对于 $\mathbf{C}^{m\times n}$，$\mathbf{C}^{n\times l}$ 以及 $\mathbf{C}^{m\times l}$ 上的同类广义矩阵范数 $\|\cdot\|$ 有：

(4) $\|\boldsymbol{AB}\|\leqslant\|\boldsymbol{A}\|\,\|\boldsymbol{B}\|$， $(\forall\boldsymbol{A}\in\mathbf{C}^{m\times n},\forall\boldsymbol{B}\in\mathbf{C}^{n\times l})$. (相容条件)

则称$\|A\|$为A的矩阵范数.

在定义 4.2.1 相容条件（4）中，当$l=1$时，B与AB分别是n维和m维列向量，由此我们可以建立一种与向量范数相联系并且相容的矩阵范数，即$\|Ax\|_\beta \leqslant \|A\|\|x\|_\alpha$

当$A\in C^{m\times n}$，且α-范数与β-范数相同时，有$\|Ax\|_\alpha \leqslant \|A\|\|x\|_\alpha$，此时称矩阵范数$\|A\|$与向量的$\alpha$-范数$\|x\|_\alpha$相容.

【例 4.2.1】 设$A\in C^{n\times n}$，规定$\|A\|=\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}}$，证明$\|A\|=\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}}$是一个矩阵范数，这个范数称为矩阵$A$的 Frobenius 范数，记作$\|A\|_F=\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}}$，且$\|A\|_F$与$\|x\|_2$相容.

证 矩阵A可以视为C^{n^2}中的向量，所以$\|A\|_F$也就是向量A的 2-范数，又设$A_i=(a_{i1},a_{i2},\cdots,a_{in})$，$x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，则

$$Ax=\begin{pmatrix}A_1\\A_2\\\vdots\\A_n\end{pmatrix}x=\begin{pmatrix}A_1x\\A_2x\\\vdots\\A_nx\end{pmatrix}$$

由 Cauchy 不等式有

$$|A_ix|^2=|a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n|^2\leqslant\sum_{j=1}^{n}|a_{ij}|^2\sum_{j=1}^{n}|x_j|^2=\|A\|_2^2\|x\|_2^2\text{，所以}$$

$$\|Ax\|_2^2=\sum_{i=1}^{n}|A_ix|^2\leqslant\left(\sum_{i=1}^{n}\|A_i\|_2^2\right)\|x\|_2^2\leqslant\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)\|x\|_2^2=\|A\|_F^2\|x\|_2^2$$

因此$\|Ax\|_2\leqslant\|A\|_F\|x\|_2$.

定理 4.2.1 $\|A\|_F=\sqrt{A^{\mathrm{H}}A\text{的所有特征值的和}}$.

证 $\|A\|_F=\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}}=\sqrt{tr(A^{\mathrm{H}}A)}=\sqrt{A^{\mathrm{H}}A\text{的所有特征值的和}}$.

下面的定理给出一种定义矩阵范数的方法，并使得确定的矩阵范数与已知的向量范数相容.

定理 4.2.2 已知线性空间C^m与C^n上的同类范数$\|\cdot\|_\alpha$，并设$A\in C^{m\times n}$，则$\|A\|=\max\limits_{x\neq0}\dfrac{\|Ax\|_\alpha}{\|x\|_\alpha}$是$C^{m\times n}$上的矩阵范数，此范数与已知的向量范数相容. 这个矩阵范数称为由向量范数$\|\cdot\|_\alpha$诱导的诱导范数（或从属范数）. 当$\|x\|_\alpha=1$时，称$\|A\|=\max\limits_{\|x\|=1}\|Ax\|_\alpha$为从属范数.

证 略.

由定理4.2.2可知，$\|A\|\geqslant\dfrac{\|Ax\|_\alpha}{\|x\|_\alpha}$，因此$\|Ax\|_\alpha\leqslant\|A\|\|x\|_\alpha$. 一般的有，$\|x\|_\alpha=\|Ex\|_\alpha\leqslant\|E\|\|x\|_\alpha$，所以，$\|E\|\geqslant1$. 但当$\|x\|_\alpha=1$时，$\|E\|=\max\limits_{\|x\|=1}\|Ex\|_\alpha=\|x\|_\alpha=1$.

由此可知，向量范数与矩阵范数存在着密切的关系. 对应向量的 1-范数，2-范数，∞-范数，就可以确定三个矩阵范数$\|A\|_1$，$\|A\|_2$和$\|A\|_\infty$.

定理 4.2.3 设$A(a_{ij})\in C^{m\times n}$，$x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in C^n$，则从属于向量的$\infty$-范数，2-范数和 1-范数的矩阵范数分别为：

(1) $\|A\|_\infty = \max\limits_i \sum\limits_{j=1}^{n} |a_{ij}|$;

(2) $\|A\|_2 = \sqrt{\lambda_1}$($\lambda_1$ 为 A^HA 的最大特征值)；

(3) $\|A\|_1 = \max\limits_j \sum\limits_{i=1}^{m} |a_{ij}|$.

分别称之为矩阵 A 的行（和）范数，谱范数与列（和）范数.

证 只就 (1)，(2) 给出证明，(3) 的证明同 (1).

(1) 设 $x=(x_1,x_2,\cdots,x_n)^T\neq 0$，不妨设 $A\neq 0$ 记

$$t=\max_{1\leqslant i\leqslant n}|x_i|,\quad \mu=\max_{1\leqslant i\leqslant n}\sum_{j=1}^{n}|a_{ij}|$$

则

$$\|Ax\|_\infty=\max_{1\leqslant i\leqslant n}\left|\sum_{j=1}^{n}a_{ij}x_j\right|\leqslant t\cdot\max_{1\leqslant i\leqslant n}\sum_{j=1}^{n}|a_{ij}|$$

这说明对任意非零 $x\in \mathbf{R}^n$，有

$$\frac{\|Ax\|_\infty}{\|x\|_\infty}\leqslant\mu$$

另一方面，设 $\mu=\sum\limits_{j=1}^{n}|a_{i_0j}|$，取向量 $x_0=(x_1,x_2,\cdots x_n)^T$，其中 $x_j=\text{sign}(a_{i_0j}),j=1,2,\cdots,n$，显然 $\|x_0\|_\infty=1$，且 Ax_0 的第 i_0 的 i_0 个分量为

$$\sum_{j=1}^{n}|a_{i_0j}x_j|=\sum_{j=1}^{n}a_{i_0j}=\mu$$

这说明 $\|Ax_0\|=\mu$，即 $\|A\|_\infty=\mu=\max\limits_{1\leqslant j\leqslant n}\sum\limits_{j=1}^{n}|a_{ij}|$

(2) 由于

$$\|A\|_2^2=(Ax)^T(Ax)=x^T(A^TA)x\geqslant 0,$$

对一切 $x\in\mathbf{R}^n$，从而 A^TA 是半正定的. 设 A^TA 的特征值为

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n\geqslant 0,$$

再设 ξ_1，ξ_2，…，ξ_n 为 A^TA 的分别对应于 λ_1，λ_2，…，λ_n 的正交规范的特征向量，则对任一向量 $x\in\mathbf{R}^n$，有

$$x=\sum_{i=1}^{n}k_i\xi_i\quad k_i\text{ 为组合系数}$$

$$\frac{\|A\|_2^2}{\|x\|_2^2}=\frac{x^T(A^TA)x}{x^Tx}=\frac{\sum\limits_{i=1}^{n}k_i^2\lambda_i}{\sum\limits_{i=1}^{n}k_i^2}\leqslant\lambda_1$$

另一方面，取 $x=\xi_1$ 则

$$\frac{\|A\|_2^2}{\|x\|_2^2}=\frac{\xi_1^T(A^TA)\xi_1}{\xi_1^T\xi_1}=\lambda_1$$

故 $\|A\|_2=\max\limits_{x\neq 0}\dfrac{\|Ax\|_2}{\|x\|_2}=\sqrt{\lambda_1}=\sqrt{\lambda_{\max}(A^TA)}$.

由定理 4.2.3 看出，计算一个矩阵的 $\|A\|_\infty$，$\|A\|_1$ 还是比较容易的，而矩阵的 2-范

数 $\|A\|_2$ 在计算上不方便，但由于它有许多好的性质. 所以，它在理论上是有用的.

定理 4.2.4　设 $A\in C^{m\times n}$，U，V 分别是 m、n 阶酉矩阵，则

$$\|UAV\|_F=\|UA\|_F=\|AV\|_F=\|A\|_F,\ \|UAV\|_2=\|UA\|_2=\|AV\|_2=\|A\|_2$$

证　略.

定理 4.2.5　在 $C^{m\times n}$上，所有满足定义 4.2.1 的矩阵范数都是相互等价的.

证　略.

定义 4.2.2　设 $A\in C^{m\times n}$，它的 n 个特征值是 λ_1，λ_2，…，λ_n，则称

$$\rho(A)=\max\{|\lambda_1|,|\lambda_2|,\cdots,|\lambda_n|\}$$

为 A 的谱半径.

定理 4.2.6　(1) 设 A 是方阵，则 $\|A\|\leqslant\rho(A)$.

(2) 若 A 是可逆矩阵，λ 是 A 的任一特征值，则有

$$\frac{1}{\|A^{-1}\|}\leqslant|\lambda|.$$

证　(1) 设 λ 是 A 的任一特征值，x 是相应的特征向量，则有 $Ax=\lambda x$，因此

$$|\lambda|\ \|x\|=\|\lambda x\|=\|Ax\|\leqslant\|A\|\ \|x\|$$

故　$|\lambda|\leqslant\|A\|$，从而 $\|A\|\leqslant\rho(A)$.

(2) λ^{-1}是 A^{-1}特征值，x 是相应的特征向量，因此

$$|\lambda|^{-1}\|x\|=\|\lambda^{-1}x\|=\|A^{-1}x\|\leqslant\|A^{-1}\|\ \|x\|$$

故 $|\lambda|^{-1}\leqslant\|A\|^{-1}$，因此有 $\dfrac{1}{\|A^{-1}\|}\leqslant|\lambda|$.

进一步，有下面结果.

定理 4.2.7　设 $\|A\|_\alpha$ 为与某种向量范数 $\|\cdot\|_\alpha$ 相容的矩阵范数，则

$$\rho(A)=\inf_\alpha\|A\|_\alpha$$

证　对任意矩阵 A，总存在相似变换化 A 为上三角形，即存在可逆矩阵 P，使

$$PAP^{-1}=\Lambda+U$$

其中，$\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$，各 λ_i 为 A 的特征值，U 是主对角元为零的上三角矩阵 $(u_{ij})_{n\times n}$.

作矩阵

$$D=\mathrm{diag}(1,\delta^{-1},\cdots,\delta^{1-n}),\quad\delta>0$$

于是

$$D(\Lambda+U)D^{-1}=\Lambda+B=C$$

其中矩阵 B 的元素为

$$b_{ij}=\begin{cases}0, & j\leqslant i\\ u_{ij}\delta^{j-i}=\delta(u_{ij}\delta^{j-i-1}), & j>i\end{cases}$$

令 $Q=DP$，则有

$$QAQ^{-1}=C$$

矩阵 A 和 C 相似，所以有相同的特征值.

今规定向量范数 $\|\cdot\|_\alpha$ 如下：令 $y=Qx$，则有

$$\|A\|_\alpha=\max_{x\neq0}\frac{\|Ax\|_\alpha}{\|x\|_\alpha}=\max_{x\neq0}\frac{\|QAx\|_2}{\|Qx\|_2}$$

$$= \max_{y \neq 0} \frac{\| Cy \|_2}{\| y \|_2} = \| C \|_2 \leqslant \| A \|_2 + \| B \|_2$$

因为 A 为对角矩阵，所以

$$\| A \|_2 = \rho(A) = \rho(A)$$

其次，存在常数 $k>0$，使

$$\| B \|_2 \leqslant \delta K$$

从而有

$$\rho \| A \|_\alpha \leqslant \rho(A) + \delta K$$

今对给定的 $\varepsilon>0$，取 $\delta=\frac{\varepsilon}{K}$，上式就化为

$$\rho(A) \leqslant \| A \|_\alpha \leqslant \rho(A) + \varepsilon$$

定理 4.2.8 如果 $A \in \mathbf{R}^{n \times n}$ 为对称矩阵，则

$$\| A \|_2 = \rho(A)$$

事实上

$$\| A \|_2^2 = \rho(A^{\mathrm{T}}A) = \rho(A^2) = (\rho(A))^2$$

由于 $\| A \|_2^2 = \sqrt{\rho\ (A^{\mathrm{T}}A)}$，所以 $\| A \|_2$ 也常记为 $\| A \|_{SP}$ 并称为谱范数.

定理 4.2.9 如果 $\| E \| =1$，$\| B \| <1$ 则 $E \pm B$ 为非奇异矩阵，且

$$\| (E \pm B)^{-1} \| \leqslant \frac{1}{1 - \| B \|}$$

其中 $\| \cdot \|$ 是指矩阵的算子范数.

下面介绍矩阵范数理论在离散线性系统稳定性判别中的应用.

$A \in \mathbf{C}^{n \times n}$，它的谱半径 $\rho(A)$ 是一个相当重要的量，在后续内容中将看到 $\rho(A)$ 的确定对于由 A 形成的幂级数 $\sum_{n=0}^{+\infty} c_k A^k$ 的收敛以及基于 A 所作的矩阵迭代的收敛起着重要的作用. 在稳定性分析中起着重要的作用. 首先，先介绍一下矩阵范数理论中与稳定性相关的定理.

定理 4.2.10 有一离散系统，$x(k+1) = Ax(k)$，$k=0, 1, 2, \cdots$对任何 $x_0 \in \mathbf{C}^n$，$x(0) = x_0$ 系统为原点平衡状态，即 $x_0=0$，当且仅当 A 的谱半径 $\rho(A)<1$ 时，则有即系统在原点平衡状渐近稳定.

该定理的证明略.

有了定理 4.2.10，我们知道只要能求出给定矩阵的谱半径就能判定系统的稳定性. 但是，求一个矩阵的特征值也不是简单的事，特别是复杂的矩阵，如果能不借助于求矩阵的特征值而能对谱半径作近似估计，则对我们对线性系统的稳定性十分有利. 于是想到用定理 4.2.11 来估计出矩阵的谱半径.

定理 4.2.11 $A \in \mathbf{C}^{n \times n}$，$\eta_1$，$\eta_2$，…，$\eta_n$ 是任意一组正实数，令

$$\rho_1 = \max_i \Big[\big(\sum_{j=1}^{n} \mid a_{ij} \mid \eta_j \big) \big/ \eta_i \Big], \rho_2 = \max_j \Big(\sum_{i=1}^{n} \frac{\mid a_{ij} \mid}{\eta_i} \eta_j \Big)$$

则 $\rho(A) = \min(\rho_1, \rho_2)$.

证 令 $Y = \mathrm{diag}[\eta_1, \eta_2, \cdots, \eta_n]$，则 $Y^{-1}AY$ 与 A 有相同的谱半径，但

$$
Y^{-1}AY = \begin{bmatrix} a_{11} & a_{12}\left(\frac{\eta_2}{\eta_1}\right) & \cdots & a_{1n}\left(\frac{\eta_n}{\eta_1}\right) \\ a_{21}\left(\frac{\eta_1}{\eta_2}\right) & a_{22} & \cdots & a_{2n}\left(\frac{\eta_n}{\eta_2}\right) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}\left(\frac{\eta_1}{\eta_n}\right) & a_{n2}\left(\frac{\eta_2}{\eta_n}\right) & \cdots & a_{nn} \end{bmatrix}
$$

而 $\|Y^{-1}AY\|_\infty=\rho_1$，$\|Y^{-1}AY\|_1=\rho_2$. 由此从 $\rho(A)=\rho(Y^{-1}AY)\leqslant\min[\|Y^{-1}AY\|_1,\|Y^{-1}AY\|_\infty]$ 可知定理成立.

【例 4.2.2】　设有一离散系统，$x(k+1)=Ax(k)$，$k=0，1，2，\cdots$，其中矩阵 $A=\begin{bmatrix}0&1&0\\0&0&1\\0&\frac{1}{3}&0\end{bmatrix}$，试用矩阵范数分析法判断系统在平衡状态 $x_e=0$ 处的渐近稳定性.

解　采用上述矩阵范数法，任取一组正实数 $\eta_1=3$，$\eta_2=2$，$\eta_3=1$ 经计算，可得

$$\rho(A)\leqslant\min(\rho1,\rho2)=2/3$$

由于 $\rho(A)\leqslant2/3<1$，所以，由定理 4.2.10 可判断该系统在平衡状态 $x_e=0$ 处渐稳定.

下面介绍矩阵范数在误差分析中的应用.

在求解线性代数方程组时，利用线性代数的知识，给出系数矩阵 A 和自由项 b 后，就可求出解 x. 实践中，A 和 b 往往是实验观测数据或是计算所得结果，A 和 b 往往有一定的误差，因此处理的线性方程组 $Ax=b$ 实际上变成了

$$(A+\delta A)(x+\delta x)=(b+\delta b)$$

δA 或 δb 与 δx 的关系怎样，是人们十分关心的问题.

【例 4.2.3】　解方程组 $Ax=b$，其中

$$A=\begin{bmatrix}1/2&1/3&1/4\\1/3&1/4&1/5\\1/4&1/5&1/6\end{bmatrix}$$

现用绝对精确的计算（即不带任何舍入误差的计算）求解，可以看出

$$x_1=72b_1-240b_2+180b_3$$
$$x_2=-240b_1+900b_2-720b_3$$
$$x_3=180b_1-720b_2+600b_3$$

此时，对于两组不同的自由项，即

$$b=(b_1,b_2,b_3)^T,\tilde{b}=(b_1-\varepsilon,b_2+\varepsilon,b_3-\varepsilon)^T$$

它的差只有 $\delta b=\tilde{b}-b=(-\varepsilon,\ \varepsilon,\ -\varepsilon)^T$，而所得解 x 与 x^T 之差却是

$$\delta x=\tilde{x}-x=(-492\varepsilon,1860\varepsilon,-1500\varepsilon)^T$$

换句话说，两组不同的右端其分量之差不过是 $|\varepsilon|$，可是解的差却高达 $|\varepsilon|$ 之 1860 倍.

对于这样的方程组，不管用什么样的数值方法，总很难（甚至不可能）算出合理的（与真正精确解相差不大的）解，像这样的方程组或矩阵 A 就叫做病态的.

定义 4.2.3　如果矩阵 A 或自由项 b 的微小变化，引起方程组 $Ax=b$ 解的巨大变化，则

称此方程组为病态方程组，矩阵 A 为病态矩阵，否则称方程组为良态方程组，A 为良态矩阵.

应该注意，矩阵的病态性质是矩阵本身的特性. 下面研究方程组 $Ax=b$，希望能找出刻画矩阵病态性质的量，为了简单，先假设 $\delta A=0$，讨论自由项对 x 的影响，再假设 $\delta b=0$，讨论系数矩阵与解 x 的关系.

如果方程组 $Ax=b$ 中系数矩阵 A 是精确的，自由项 b 有误差 δb，相应的解为（$x+\delta x$），则

$$A(x+\delta x)=(b+\delta b)$$

利用关系式 $Ax=b$ 有

$$A\delta x=\delta b \text{ 即 } \delta x=A^{-1}\delta b$$

两边取范数，有

$$\frac{\|\delta x\|}{\|x\|}\geqslant\frac{\|\delta b\|}{\|A\|\ \|x\|}\geqslant\frac{\|\delta b\|}{\|A\|\ \|A^{-1}\|\ \|b\|}$$

$$\frac{\|\delta x\|}{\|x\|}\leqslant\frac{\|A^{-1}\|\ \|\delta b\|}{\|x\|}=\frac{\|A\|\ \|A^{-1}\|\ \|\delta b\|}{\|A\|\ \|x\|}\leqslant\frac{\|A\|\ \|A^{-1}\|\ \|\delta b\|}{\|b\|}$$

故

$$\frac{1}{\|A\|\ \|A^{-1}\|}\ \frac{\|\delta b\|}{\|b\|}\leqslant\frac{\|\delta x\|}{\|x\|}\leqslant\|A\|\ \|A^{-1}\|\ \frac{\|\delta b\|}{\|b\|}$$

如果 b 是精确的，A 有微小误差 δA，相应的解为 $x+x\delta$，则

$$(A+\delta A)(x+x\delta)=b$$

这种情况比前一种情况复杂，$A+\delta A$ 可能是奇异矩阵. 为了简单，我们要求 δA 能够保证使 $A+\delta A$ 非奇异，因为

$$A+\delta A=A(E+A^{-1}\delta A)$$

由定理 4.2.12，当 $\|A^{-1}\delta A\|<1$ 时，$E+A^{-1}\delta A$ 非奇异，从而 $A+\delta A$ 非奇异，且

$$\|(A+\delta A)^{-1}\|=\|A^{-1}\|\ \|(E+A^{-1}\delta A)^{-1}\|\leqslant\frac{\|A^{-1}\|}{1-\|A^{-1}\delta A\|}$$

下面的讨论使用了更强的条件，即 $\|A^{-1}\|\ \|\delta A\|<1$，于是

$$\|\delta x\|\leqslant\frac{\|A^{-1}\|\ \|\delta A\|\ \|x\|}{1-\|A^{-1}\delta A\|}\leqslant\frac{\|A^{-1}\|\ \|\delta A\|\ \|x\|}{1-\|A^{-1}\|\ \|\delta A\|}$$

即

$$\frac{\|\delta x\|}{\|x\|}\leqslant\frac{\|A^{-1}\|\ \|A\|\ \dfrac{\|\delta A\|}{\|A\|}}{1-\|A^{-1}\|\ \|A\|\ \dfrac{\|\delta A\|}{\|A\|}}$$

另一方面，利用 $Ax=b$ 有

$$\delta x=-A^{-1}\delta A(x+\delta x)$$

两边取范数，有

$$\frac{\|\delta x\|}{\|x+\delta x\|}\leqslant\|A^{-1}\|\ \|\delta A\|=\|A^{-1}\|\ \|A\|\ \frac{\|\delta A\|}{\|A\|}$$

由此可以看出，量 $\|A\|\ \|A^{-1}\|$ 越小由 A 或 b 的相对误差引起的解的相对误差就越小，量 $\|A\|\ \|A^{-1}\|$ 越大，解的相对误差就可能越大. 所以量 $\|A\|\ \|A^{-1}\|$ 实际上刻画了

解对原始数据变化的灵敏程度，刻画了方程组的病态程度．于是引进下述定义．

定义 4.2.4　设 $\boldsymbol{A}$ 为非奇异矩阵，称数

$$\operatorname{cond}(\boldsymbol{A}) = \|\boldsymbol{A}\| \|\boldsymbol{A}^{-1}\|$$

为矩阵 $\boldsymbol{A}$ 的条件数．

矩阵的条件数与范数有关，常使用的条件数有

(1) $\operatorname{cond}_\infty(\boldsymbol{A}) = \|\boldsymbol{A}\|_\infty \|\boldsymbol{A}^{-1}\|_\infty$；

(2) $\operatorname{cond}_2(\boldsymbol{A}) = \|\boldsymbol{A}\|_2 \|\boldsymbol{A}^{-1}\|_2$；

(3) $\operatorname{cond}_1(\boldsymbol{A}) = \|\boldsymbol{A}\|_1 \|\boldsymbol{A}^{-1}\|_1$.

当 $\boldsymbol{A}$ 为对称矩阵时

$$\operatorname{cond}_2(\boldsymbol{A}) = \frac{|\lambda_1|}{|\lambda_n|}$$

其中 λ_1，λ_n 分别为 $\boldsymbol{A}$ 的绝对值最大和最小的特征值．

对病态的方程组的进一步讨论，大家可阅读数值分析的相关内容．

习　题　4

4.1　设矩阵 $\boldsymbol{A}\in \mathbf{C}^{m\times n}$ 列满秩，给定 $\mathbf{C}^m$ 中的一种向量范数 $\|\cdot\|$，证明：

$$\|\boldsymbol{x}\| = \|\boldsymbol{Ax}\|, \quad \forall x \in \mathbf{C}^n$$

是 $\mathbf{C}^n$ 中的向量范数．

4.2　证明：$\forall \boldsymbol{x}\in\mathbf{C}^n$，有

$$\|\boldsymbol{x}\|_\infty \leqslant \|\boldsymbol{x}\|_2 \leqslant \|\boldsymbol{x}\|_1$$

4.3　设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，且 $\|\boldsymbol{A}\|<1$，证明：

(1) $\boldsymbol{E}-\boldsymbol{A}$ 可逆；

(2) 若 $\|\boldsymbol{E}\|=1$，则 $\|(\boldsymbol{E}-\boldsymbol{A})^{-1}\| \leqslant \dfrac{1}{1-\|\boldsymbol{A}\|}$；

(3) 若 $\|\boldsymbol{E}\|=1$，$\|\boldsymbol{E}-(\boldsymbol{E}-\boldsymbol{A})^{-1}\| \leqslant \dfrac{\|\boldsymbol{A}\|}{1-\|\boldsymbol{A}\|}$.

4.4　设单位向量 $\boldsymbol{y}\in\mathbf{R}^n(n>1)$，令 $\boldsymbol{B}=\boldsymbol{E}-\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}$，证明：

(1) $\|\boldsymbol{B}\|_2=1$；

(2) 对于 $\boldsymbol{x}\in\mathbf{R}^n$，若 $\boldsymbol{Bx}\neq\boldsymbol{x}$，则 $\|\boldsymbol{Bx}\|_2\leqslant\|\boldsymbol{x}\|$.

4.5　设 $\boldsymbol{A}=\begin{pmatrix}1&2&3\\2&3&4\\3&4&5\end{pmatrix}$，求 $\|\boldsymbol{A}\|_1$、$\|\boldsymbol{A}\|_2$ 和 $\|\boldsymbol{A}\|_F$.

4.6　设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$ 的特征值为 $\lambda_1,\lambda,\cdots\lambda_n$，证明：

$$\|\boldsymbol{A}\|_F \geqslant (|\lambda_1|^2+|\lambda_2|^2+\cdots|\lambda_n|^2)^{1/2}$$

4.7　设 n 阶实对称阵 $\boldsymbol{A}$ 满足 $(\boldsymbol{A}+\boldsymbol{E})(\boldsymbol{A}-2\boldsymbol{E})^2=0$，$\operatorname{rank}(\boldsymbol{A}+\boldsymbol{E})=r$，求 $\|\boldsymbol{A}\|_2$ 和 $\|\boldsymbol{A}\|_F$.

设 $\boldsymbol{A}$ 是 n 阶可逆矩阵，$\boldsymbol{B}$ 是 n 阶矩阵．若对某种矩阵范数有 $\|\boldsymbol{B}\|<\dfrac{1}{\|\boldsymbol{A}^{-1}\|}$，则 $\boldsymbol{A}+\boldsymbol{B}$ 可逆．

4.8　设 $\boldsymbol{A}=(a_{ij})_{n\times n}\in\mathbf{C}^{n\times n}$，定义实数

$$\|A\|_G = n\max_{i,j}|a_{i,j}|$$

证明：$\|A\|_G$ 是 $C^{n\times n}$ 中的矩阵范数，且与 C^n 中向量的 2-范数相容.

4.9　举例说明 $C^{n\times n}(n>1)$ 中的矩阵范数 $\|A\|_1$ 与 C^n 中向量的 2-范数不相容.

4.10　设 A，$B\in C^{n\times n}$ 都是可逆矩阵，且齐次线性方程组 $(A+B)x=0$ 有非零解. 证明：对于 $C^{n\times n}$ 中的任何矩阵阵范数 $\|\cdot\|$ 都有 $\|A^{-1}B\|\geqslant 1$ 及 $\|B^{-1}A\|\geqslant 1$.

4.11　已知 $A=\begin{pmatrix}2 & 1\\ 1 & 3\end{pmatrix}$，$\delta A=\begin{pmatrix}0 & 0.2\\ 0.4 & -0.3\end{pmatrix}$，试估计 $\dfrac{\|A^{-1}-(A+\delta A)^{-1}\|_\infty}{\|A^{-1}\|_\infty}$.

第五章 矩 阵 分 析

在线性代数中，主要讨论了矩阵的代数运算，没有涉及本章将要介绍的矩阵分析理论. 矩阵分析理论的建立，同数学分析一样，也是以极限理论为基础. 它也是研究数值方法和其他数学分支以及许多工程问题的重要工具. 本章首先讨论矩阵序列的极限运算，然后介绍矩阵级数的收敛定理、函数矩阵的微分和积分. 此外，我们还将介绍矩阵的 Kronecker 积和拉直.

§5.1 矩阵序列的极限

定义 5.1.1 设有矩阵序列 $\{\boldsymbol{A}^{(k)}\}$，其中 $\boldsymbol{A}^{(k)}=(a_{ij}^{(k)})\in\mathbf{C}^{m\times n}$. 若对所有 $i=1, 2, \cdots, m$ 和 $j=1, 2, \cdots, n$，$\lim\limits_{k\to\infty}a_{ij}^{(k)}$ 存在，且 $\lim\limits_{k\to\infty}a_{ij}^{(k)}=a_{ij}$，则称 $\{\boldsymbol{A}^{(k)}\}$ 的极限存在，并称矩阵 $\boldsymbol{A}=(a_{ij})$ 为 $\{\boldsymbol{A}^{(k)}\}$ 的极限，或称 $\{\boldsymbol{A}^{(k)}\}$ 收敛于 $\boldsymbol{A}$，记为

$$\lim_{k\to\infty}\boldsymbol{A}^{(k)}=\boldsymbol{A},\text{或}\ \boldsymbol{A}^{(k)}\to\boldsymbol{A}(k\to\infty)$$

不收敛的矩阵序列称为发散.

定义 5.1.2 设有矩阵序列 $\{\boldsymbol{A}^{(k)}\}$，其中 $\boldsymbol{A}^{(k)}=(a_{ij}^{(k)})\in\mathbf{C}^{m\times n}$. 若存在 $M>0$，使得对一切 k，都有

$$|a_{ij}^{(k)}|<M,\quad (i=1,2,\cdots,m;j=1,2,\cdots,n)$$

则称矩阵序列 $\{\boldsymbol{A}^{(k)}\}$ 是有界的.

矩阵序列收敛的性质，有许多与数列收敛的性质类似.

定理 5.1.1 收敛的矩阵序列 $\{\boldsymbol{A}^{(k)}\}$ 一定有界.

证 设矩阵序列 $\{\boldsymbol{A}^{(k)}\}$ 是收敛的序列，不妨设 $\lim\limits_{k\to\infty}\boldsymbol{A}^{(k)}=\boldsymbol{A}$，对 $\varepsilon=1$，存在 $K>0$，使得对一切 $k>K$ 都有

$$|a_{ij}^{(k)}-a_{ij}|<1\quad \text{对所有}\ i=1,2,\cdots,m;j=1,2,\cdots,n$$

因此

$$|a_{ij}^{(k)}|<\max\{|a_{ij}|+1,i=1,2,\cdots,m;j=1,2,\cdots n\}$$

对所有 $i=1,2,\cdots,m$；$j=1,2,\cdots,n$ 成立. 从而矩阵序列 $\{\boldsymbol{A}^{(k)}\}$ 有界.

定理 5.1.2 设 $\boldsymbol{A}^{(k)}$，$\boldsymbol{A}\in\mathbf{C}^{m\times n}$，则

(1) $\lim\limits_{k\to\infty}\boldsymbol{A}^{(k)}=0$ 的充分且必要条件是 $\lim\limits_{k\to\infty}\|\boldsymbol{A}^{(k)}\|=\mathbf{0}$；

(2) $\lim\limits_{k\to\infty}\boldsymbol{A}^{(k)}=\boldsymbol{A}$ 的充分且必要条件是 $\lim\limits_{k\to\infty}\|\boldsymbol{A}^{(k)}-\boldsymbol{A}\|=0$.

其中 $\|\cdot\|$ 是 $\boldsymbol{A}$ 的任意一种范数.

证 (1) 由第四章知道，$\mathbf{C}^{m\times n}$ 中的范数是等价的. 因此只需对矩阵的 F-范数证明即可

$$\lim_{k\to\infty}\boldsymbol{A}^{(k)}=\mathbf{0}\Leftrightarrow\lim_{k\to\infty}a_{ij}^{(k)}=\mathbf{0}(\text{对所有}\ i=1,2,\cdots,m;j=1,2,\cdots,n)$$

$$\Leftrightarrow\lim_{k\to\infty}\Big(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}^{(k)}|^2\Big)^{\frac{1}{2}}=0$$

$$\Leftrightarrow \lim_{k\to\infty} \| A^{(k)} \|_F = 0.$$

(2) 由于$\lim\limits_{k\to\infty} A^{(k)}=A$的充分且必要条件是$\lim\limits_{k\to\infty}(A^{(k)}-A)=0$，因此利用（1）的结论即可.

定理 5.1.3 设$\lim\limits_{k\to\infty} A^{(k)}=A$，$\lim\limits_{k\to\infty} B^{(k)}=B$，则

$$\lim_{k\to\infty}(aA^{(k)}+bB^{(k)})=aA+bB$$

证 略.

定理 5.1.4 设$\lim\limits_{k\to\infty} A^{(k)}=A\in C^{m\times n}$，$\lim\limits_{k\to\infty} B^{(k)}=B\in C^{n\times s}$，则

$$\lim_{k\to\infty} A^{(k)}B^{(k)}=AB$$

证 因为$\lim\limits_{k\to\infty} A^{(k)}=A\in C^{m\times n}$，由定理 5.1.1，存在$M>0$，使得对一切$k$，都有

$$|a_{ij}^{(k)}|<M,\quad (i=1,2,\cdots,m;j=1,2,\cdots,n)$$

因此

$$\begin{aligned}\| A^{(k)}B^{(k)}-AB \|_F &= \| A^{(k)}B^{(k)}-A^{(k)}B+A^{(k)}B-AB \|_F \\ &\leqslant \| A^{(k)}B^{(k)}-A^{(k)}B \|_F + \| A^{(k)}B-AB \|_F \\ &\leqslant \| A^{(k)} \|_F \| B^{(k)}-B \|_F + \| A^{(k)}-A \|_F \| B \|_F \\ &\leqslant mnM \| B^{(k)}-B \|_F + \| A^{(k)}-A \|_F \| B \|_F.\end{aligned}$$

故

$$\lim_{k\to\infty} \| A^{(k)}B^{(k)}-AB \|_F \leqslant mnM \lim_{k\to\infty} \| B^{(k)}-B \|_F + \lim_{k\to\infty} \| A^{(k)}-A \|_F \| B \|_F = 0$$

从而 $\lim\limits_{k\to\infty} \| A^{(k)}B^{(k)}-AB \|_F=0$，即$\lim\limits_{k\to\infty} A^{(k)}B^{(k)}=AB$.

定理 5.1.5 设$A^{(k)}$，A都是可逆矩阵，$\lim\limits_{k\to\infty} A^{(k)}=A$，则$\lim\limits_{k\to\infty}(A^{(k)})^{-1}=A^{-1}$.

证 由于$A^{(k)}$，A都是可逆矩阵，而$\lim\limits_{k\to\infty} A^{(k)}=A$，因此

$$\lim_{k\to\infty}\min\{|\lambda_i^{(k)}|,i=1,2,\cdots,n\}=\min\{|\lambda_i|,i=1,2,\cdots,n\}$$

由极限定义知，对$\varepsilon=\frac{1}{2}\min\{|\lambda_i|,\ i=1,\ 2,\ \cdots,\ n\}$，存在$K>0$，当$k>K$时，有

$$\min\{|\lambda_i^{(k)}|,i=1,2,\cdots,n\}\geqslant \frac{1}{2}\min\{|\lambda_i|,i=1,2,\cdots,n\}$$

因此，存在$M>0$，使得对一切k有

$$\| [A^{(k)}]^{-1} \|_2 \leqslant M$$

故

$$\begin{aligned}\| [A^{(k)}]^{-1}A-E \|_2 &= \| [A^{(k)}]^{-1}A-[A^{(k)}]^{-1}[A^{(k)}] \|_2 \\ &\leqslant \| [A^{(k)}]^{-1} \|_2 \| A^{(k)}-A \|_2 \leqslant M \| A^{(k)}-A \|_2\end{aligned}$$

所以

$$\lim_{k\to\infty} \| [A^{(k)}]^{-1}A-E \|_2=0$$

即 $\lim\limits_{k\to\infty}[A^{(k)}]^{-1}A=E$. 由于$A^{-1}$唯一存在，因此$\lim\limits_{k\to\infty}(A^{(k)})^{-1}=A^{-1}$.

下面我们来讨论序列$\{A^{(k)}\}$的收敛性问题.

定理 5.1.6 设$A\in C^{n\times n}$. $\lim\limits_{k\to\infty} A^k=0$的充分必要条件是$\rho(A)<1$.

证 设A的 Jordan 标准形为J，则存在可逆矩阵P，使得$A=PJP^{-1}$，于是

$$A^k = PJ^kP^{-1}$$

注意到

$$J^k = \mathrm{diag}(J_1^k, J_2^k, \cdots, J_s^k)$$

$$J_i^k = \begin{bmatrix} \lambda_i^k & C_k^1\lambda_i^{k-1} & C_k^2\lambda_i^{k-2} & \cdots & C_k^{m_i-2}\lambda_i^{k-m_i} & C_k^{m_i-1}\lambda_i^{k-m_i+1} \\ 0 & \lambda_i^k & C_k^1\lambda_i^{k-1} & \cdots & C_k^{m_i-3}\lambda_i^{k-m_i-1} & C_k^{m_i-2}\lambda_i^{k-m_i} \\ 0 & 0 & \lambda_i^k & \cdots & C_k^{m_i-4}\lambda_i^{k-m_i-2} & C_k^{m_i-3}\lambda_i^{k-m_i-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i^k & C_k^1\lambda_i^{k-1} \\ 0 & 0 & 0 & \cdots & \cdots & \lambda_i^k \end{bmatrix}$$

其中 J_i 是 A 的以 λ_i 为特征值的 Jordan 块.

因此 $\lim\limits_{k\to\infty} A^k = 0$ 的充分必要条件为 $\lim\limits_{k\to\infty} J^k = 0$；

充分必要条件 $\lim\limits_{k\to\infty} J_i^k = 0$ 对 $i=1, 2, \cdots, s$ 都成立；

充分必要条件 $|\lambda_i|<1$ 对 $i=1, 2, \cdots, s$ 都成立；

充分必要条件是 $\rho(A)<1$.

推论 5.1.1 设 $A\in C^{n\times n}$. 若对某一范数 $\|\cdot\|$，$\|A\|<1$，则 $\lim\limits_{k\to\infty} A^k = 0$.

如 $A=\begin{bmatrix} 0.25 & 0.4 & -0.4 \\ -0.35 & 0.2 & 0.3 \\ 0.15 & 0.1 & 0 \end{bmatrix}$. 由于 $\|A\|_1=0.75<1$，因此 $\lim\limits_{k\to\infty} A^k = 0$.

§5.2 矩 阵 级 数

定义 5.2.1 设有矩阵序列 $\{A^{(k)}\}$，其中 $A^{(k)}=(a_{ij}^{(k)})\in C^{m\times n}$. 称

$$A^{(0)}+A^{(1)}+\cdots+A^{(k)}+\cdots$$

为矩阵级数，记为 $\sum\limits_{k=0}^{\infty} A^{(k)}$.

定义 5.2.2 称 $S^{(N)}=\sum\limits_{k=0}^{N} A^{(k)}$ 为矩阵级数 $\sum\limits_{k=0}^{\infty} A^{(k)}$ 的前 N 项部分和. 若 $\lim\limits_{k\to\infty} S^{(N)}$ 存在，且 $\lim\limits_{N\to\infty} S^{(N)}=S$，则称矩阵级数 $\sum\limits_{k=0}^{\infty} A^{(k)}$ 收敛，而且有和 S，记为 $\sum\limits_{k=0}^{\infty} A^{(k)}=S$. 若 $\lim\limits_{k\to\infty} S^{(N)}$ 不存在，则称矩阵级数 $\sum\limits_{k=0}^{\infty} A^{(k)}$ 发散.

由定义 5.2.2 知，设 $A^{(k)}=(a_{ij}^{(k)})_{m\times n}$，矩阵级数 $\sum\limits_{k=0}^{\infty} A^{(k)}$ 收敛的充分必要条件是 $m\times n$ 个数项级数 $\sum\limits_{k=0}^{\infty} a_{ij}^{(k)}$ 收敛.

定义 5.2.3 设 $A^{(k)}=(a_{ij}^{(k)})_{m\times n}$. 对矩阵级数 $\sum\limits_{k=0}^{\infty} A^{(k)}$，若 $m\times n$ 个数项级数 $\sum\limits_{k=0}^{\infty} a_{ij}^{(k)}$ 绝对收敛，则称矩阵级数 $\sum\limits_{k=0}^{\infty} A^{(k)}$ 绝对收敛.

由数项级数的性质和矩阵级数绝对收敛的定义，我们有如下定理.

定理 5.2.1 若矩阵级数 $\sum_{k=0}^{\infty}A^{(k)}$ 绝对收敛，则它一定收敛，并且任意调换各项的次序所得的新级数仍收敛，且其和不变.

定理 5.2.2 矩阵级数 $\sum_{k=0}^{\infty}A^{(k)}$ 绝对收敛的充分必要条件是对于任意一种矩阵范数 $\|\cdot\|$，$\sum_{k=0}^{\infty}\|A^{(k)}\|$ 收敛.

证 设 $A^{(k)}=(a_{ij}^{(k)})_{m\times n}$.

(1) 必要性. 由于矩阵级数 $\sum_{k=0}^{\infty}A^{(k)}$ 绝对收敛，因此 $m\times n$ 个数项级数 $\sum_{k=0}^{\infty}a_{ij}^{(k)}$ 绝对收敛. 因此存在 $M>0$，使得对一切 K，都有

$$\sum_{k=1}^{K}|a_{ij}^{(k)}|<M,\quad i=1,2,\cdots,m;j=1,2,\cdots,n$$

故

$$\sum_{k=0}^{K}\|A^{(k)}\|_F=\sum_{k=1}^{K}\left(\sum_{j=1}^{n}\sum_{i=1}^{m}|a_{ij}^{(k)}|^2\right)^{\frac{1}{2}}\leqslant mnM$$

因此 $\sum_{k=0}^{\infty}\|A^{(k)}\|_F$ 收敛. 利用范数的等价性知，$\sum_{k=0}^{\infty}\|A^{(k)}\|$ 收敛.

(2) 充分性. 对于任意一种矩阵范数 $\|\cdot\|$，$\sum_{k=0}^{\infty}\|A^{(k)}\|$ 收敛. 特别地，$\sum_{k=0}^{\infty}\|A^{(k)}\|_1$ 收敛. 因此

$$\sum_{k=0}^{\infty}|a_{ij}^{(k)}|\leqslant\sum_{k=0}^{\infty}\|A^{(k)}\|_1,\quad(\text{对所有 } i=1,2,\cdots,m;j=1,2,\cdots,n \text{ 成立})$$

故矩阵级数 $\sum_{k=0}^{\infty}A^{(k)}$ 绝对收敛.

定理 5.2.3 若矩阵级数 $\sum_{k=0}^{\infty}A^{(k)}$ 收敛（或绝对收敛），其和为 S，则 $\sum_{k=0}^{\infty}PA^{(k)}Q$ 收敛（或绝对收敛），其和为 PSQ.

证 若矩阵级数 $\sum_{k=0}^{\infty}A^{(k)}$ 收敛（或绝对收敛），其和为 S，则有

$$\lim_{N\to\infty}\left\|\sum_{k=0}^{N}A^{(k)}-S\right\|=0$$

因此

$$\lim_{N\to\infty}\left\|\sum_{k=0}^{N}PA^{(k)}Q-PSQ\right\|\leqslant\|P\|\lim_{N\to\infty}\left\|\sum_{k=0}^{N}A^{(k)}-S\right\|\|Q\|=0$$

故 $\sum_{k=0}^{\infty}PA^{(k)}Q$ 收敛，其和为 PSQ.

同理可证，若矩阵级数 $\sum_{k=0}^{\infty}A^{(k)}$ 绝对收敛，则 $\sum_{k=0}^{\infty}PA^{(k)}Q$ 绝对收敛.

§5.3 矩阵的 Kronecker 积

矩阵的 Kronecker 积（直积）是一种重要的矩阵乘积，它不仅在矩阵理论的研究中有着广泛的应用，而且在诸如信号处理与系统理论中的随机静态分析与随机向量过程分析等工程领域中也是一种基本的数学工具. 本节中，我们将介绍 Kronecker 积的基本性质，并应用它求解线性矩阵方程和矩阵微分方程.

我们知道，在线性代数中曾定义了两个矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的乘积 $\boldsymbol{AB}$，它要求 $\boldsymbol{A}$ 的列数等于 $\boldsymbol{B}$ 的行数，否则 $\boldsymbol{AB}$ 没有意义. 下面我们引入一种新的矩阵乘积运算，它对矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 没有要求.

定义 5.3.1 设 $\boldsymbol{A}=(a_{ij})_{m\times n}\in\mathbf{C}^{m\times n}$，$\boldsymbol{B}=(b_{ij})_{p\times q}\in\mathbf{C}^{p\times q}$，则称分块矩阵

$$\begin{pmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{pmatrix}$$

为矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的 Kronecker 积（直积），记为 $\boldsymbol{A}\otimes\boldsymbol{B}$.

【例 5.3.1】 设 $\boldsymbol{A}=\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$，$\boldsymbol{B}=\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$，则有

$$\boldsymbol{A}\otimes\boldsymbol{B}=\begin{pmatrix} 1 & 2 & 3 & 2 & 4 & 6 \\ 4 & 5 & 6 & 8 & 10 & 12 \\ 3 & 6 & 9 & 4 & 8 & 12 \\ 12 & 15 & 18 & 16 & 20 & 24 \end{pmatrix}$$

定理 5.3.1 只要运算可行，则有：

(1) $k(\boldsymbol{A}\otimes\boldsymbol{B})=(k\boldsymbol{A})\otimes\boldsymbol{B}=\boldsymbol{A}\otimes(k\boldsymbol{B})$；

(2) $(\boldsymbol{A}_1+\boldsymbol{A}_2)\otimes\boldsymbol{B}=\boldsymbol{A}_1\otimes\boldsymbol{B}+\boldsymbol{A}_2\otimes\boldsymbol{B}$；

(3) $\boldsymbol{A}\otimes(\boldsymbol{B}_1+\boldsymbol{B}_2)=\boldsymbol{A}\otimes\boldsymbol{B}_1+\boldsymbol{A}\otimes\boldsymbol{B}_2$；

(4) $(\boldsymbol{A}_1\otimes\boldsymbol{A}_2)\otimes\boldsymbol{A}_3=\boldsymbol{A}_1\otimes(\boldsymbol{A}_2\otimes\boldsymbol{A}_3)$；

(5) $(\boldsymbol{A}_1\otimes\boldsymbol{B}_1)(\boldsymbol{A}_2\otimes\boldsymbol{B}_2)=(\boldsymbol{A}_1\boldsymbol{A}_2)\otimes(\boldsymbol{B}_1\boldsymbol{B}_2)$；

(6) $(\boldsymbol{A}\otimes\boldsymbol{B})^{-1}=\boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1}$；

(7) $(\boldsymbol{A}\otimes\boldsymbol{B})^{H}=\boldsymbol{A}^{H}\times\boldsymbol{B}^{H}$；

(8) 两个酉（正交）矩阵的 Kronecker 积仍是酉（正交）矩阵.

证 我们仅证明 (5)，其余的结论，请读者自行证明. 设 $\boldsymbol{A}_1=(a_{ij})_{m\times n}$，$\boldsymbol{A}_2=(b_{ij})_{n\times s}$，$\boldsymbol{B}_1$，$\boldsymbol{B}_2$ 可乘，则

$$(\boldsymbol{A}_1\otimes\boldsymbol{B}_1)(\boldsymbol{A}_2\otimes\boldsymbol{B}_2)=\begin{pmatrix} a_{11}\boldsymbol{B}_1 & a_{12}\boldsymbol{B}_1 & \cdots & a_{1n}\boldsymbol{B}_1 \\ a_{21}\boldsymbol{B}_1 & a_{22}\boldsymbol{B}_1 & \cdots & a_{2n}\boldsymbol{B}_1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B}_1 & a_{m2}\boldsymbol{B}_1 & \cdots & a_{mn}\boldsymbol{B}_1 \end{pmatrix}\begin{pmatrix} b_{11}\boldsymbol{B}_2 & b_{12}\boldsymbol{B}_2 & \cdots & b_{1s}\boldsymbol{B}_2 \\ b_{21}\boldsymbol{B}_2 & b_{22}\boldsymbol{B}_2 & \cdots & b_{2s}\boldsymbol{B}_2 \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1}\boldsymbol{B}_2 & b_{n2}\boldsymbol{B}_2 & \cdots & b_{ns}\boldsymbol{B}_2 \end{pmatrix}$$

$$
=\begin{pmatrix}\sum_{i=1}^{n}a_{1i}b_{i1}B_1B_2 & \sum_{i=1}^{n}a_{1i}b_{i2}B_1B_2 & \cdots & \sum_{i=1}^{n}a_{1i}b_{is}B_1B_2\\ \sum_{i=1}^{n}a_{2i}b_{i1}B_1B_2 & \sum_{i=1}^{n}a_{2i}b_{i2}B_1B_2 & \cdots & \sum_{i=1}^{n}a_{2i}b_{is}B_1B_2\\ \vdots & \vdots & \ddots & \vdots\\ \sum_{i=1}^{n}a_{mi}b_{i1}B_1B_2 & \sum_{i=1}^{n}a_{mi}b_{i2}B_1B_2 & \cdots & \sum_{i=1}^{n}a_{mi}b_{is}B_1B_2\end{pmatrix}=(A_1A_2)\otimes(B_1B_2)
$$

定理 5.3.2 设 $A\in C^{m\times n}$，$B\in C^{p\times q}$，则 $\mathrm{rank}(A\otimes B)=\mathrm{rank}(A)\cdot\mathrm{rank}(B)$

证 设 $\mathrm{rank}(A)=r$，则存在非奇异阵 P，Q 使得 $A=P\begin{pmatrix}E_r & 0\\ 0 & 0\end{pmatrix}Q$. 因此

$$
A\otimes B=\left[P\begin{pmatrix}E_r & 0\\ 0 & 0\end{pmatrix}Q\right]\otimes(E_pBE_q)=(P\otimes E_p)\begin{pmatrix}E_r\times B & 0\\ 0 & 0\end{pmatrix}_{mp\times np}(Q\otimes E_q)
$$

由定理 5.3.1（6）知，$P\otimes E_p$，$Q\times E_q$ 非奇异，注意到 $E_r\otimes B=\mathrm{diag}(B, B, \cdots, B)$，因此

$$
\mathrm{rank}(A\otimes B)=\mathrm{rank}\begin{pmatrix}E_r\otimes B & 0\\ 0 & 0\end{pmatrix}=r\cdot\mathrm{rank}(B)=\mathrm{rank}(A)\cdot\mathrm{rank}(B).
$$

定理 5.3.3 设 $A\in C^{m\times m}$，$B\in C^{n\times n}$，则

(1) $|A\otimes B|=|A|^n|B|^m$；

(2) $\mathrm{tr}(A\otimes B)=\mathrm{tr}(A)\cdot\mathrm{tr}(B)$.

证 设 λ_1，λ_2，…，λ_m 是 A 的特征值，J 是 A 的 Jordan 标准形. 则存在非奇异阵 P，使得 $A=PJP^{-1}$. 因此

$$
A\otimes B=(PJP^{-1})\otimes(E_nBE_n)=(P\otimes E_n)(J\otimes B)(P^{-1}\otimes E_n)
$$

$$
|A\otimes B|=|J\otimes B|,\mathrm{tr}(A\otimes B)=\mathrm{tr}(J\otimes B)
$$

注意到 $J\otimes B$ 是分块上三角阵，其对角线上的矩阵是 λ_1B，λ_2B，…，λ_mB，因此：

(1) $|J\otimes B|=|\lambda_1B||\lambda_2B|\cdots|\lambda_mB|=(\lambda_1\lambda_2\cdots\lambda_m)^n|B|^m=|A|^n|B|^m$.

故 $|A\otimes B|=|A|^n|B|^m$.

(2) $\mathrm{tr}(J\otimes B)=\lambda_1\mathrm{tr}(B)+\lambda_2\mathrm{tr}(B)+\cdots+\lambda_m\mathrm{tr}(B)=\mathrm{tr}(A)\cdot\mathrm{tr}(B)$.

故 $\mathrm{tr}(A\otimes B)=\mathrm{tr}(A)\cdot\mathrm{tr}(B)$.

定义 5.3.2 设 $A\in C^{m\times m}$，$B\in C^{n\times n}$，称

$$
A\otimes E_n+E_m\otimes B^{\mathrm{T}}
$$

为矩阵 A 与 B 的 Kronecker 和.

定理 5.3.4 设 $A\in C^{m\times m}$，$B\in C^{n\times n}$. 记 $\Lambda(A)=\{\lambda\mid\lambda$ 是 A 的特征值$\}$. 则有：

(1) $\Lambda(A\otimes B)=\{\lambda u\mid\lambda\in\Lambda(A),u\in\Lambda(B)\}$；

(2) $\Lambda(A\otimes E_n+E_m\otimes B^{\mathrm{T}})=\{\lambda+u\mid\lambda\in\Lambda(A),u\in\Lambda(B)\}$.

证 (1) 设 λ_1，λ_2，…，λ_m 是 A 的特征值，J_A 是 A 的 Jordan 标准形，u_1，u_2，…，u_n 是 B 的特征值，J_B 是 B 的 Jordan 标准形. 则存在非奇异阵 $P\in C^{m\times m}$，$Q\in C^{n\times n}$，使得

$$
A=PJ_AP^{-1},B=QJ_BQ^{-1}
$$

因此

$$
A\otimes B=(PJ_AP^{-1})\otimes(QJ_BQ^{-1})
$$

$$= (P\otimes Q)(J_A\otimes J_B)(P^{-1}\otimes Q^{-1})$$
$$= (P\otimes Q)(J_A\otimes J_B)(P\otimes Q)^{-1}$$

则

$$\Lambda(A\otimes B) = \Lambda(J_A\otimes J_B) = \{\lambda_i u_j \mid i=1,2,\cdots,m; j=1,2,\cdots,n\}$$

(2) 设 J_A 是 A 的 Jordan 标准形，λ_1，λ_2，…，λ_m 是 A 的特征值，u_1，u_2，…，u_n 是 B 的特征值，J_{B^T} 是 B^T 的 Jordan 标准形. 则存在非奇异阵 $P\in C^{m\times m}$，$Q\in C^{n\times n}$，使得

$$A = PJ_AP^{-1}, B^T = QJ_{B^T}Q^{-1}$$

因此

$$A\otimes E_n + E_m\otimes B^T = (PJ_AP^{-1})\otimes(QE_nQ^{-1}) + (PE_mP^{-1})\otimes(QJ_{B^T}Q^{-1})$$
$$= (P\otimes Q)(J_A\otimes E_n + E_m\otimes J_{B^T})(P^{-1}\otimes Q^{-1})$$
$$= (P\otimes Q)(J_A\otimes E_n + E_m\otimes J_{B^T})(P\otimes Q)^{-1}.$$

则 $A\otimes E_n+E_m\otimes B^T$ 与 $J_A\otimes E_n+E_m\otimes J_{B^T}$ 有相同的特征值. 注意 $J_A\otimes E_n+E_m\otimes J_{B^T}$ 是对角线的下方全为零的矩阵，对角线上的元素为

$$\lambda_i + u_j, i=1,2,\cdots,m; j=1,2,\cdots,n$$

由此有 $\Lambda(A\otimes E_n + E_m\otimes B^T) = \{\lambda + u \mid \lambda\in\Lambda(A), u\in\Lambda(B)\}$.

推论 5.3.1 设 x 是 A 对应于特征值 λ 的特征向量，y 是 B 对应于特征值 u 的特征向量，则 $x\otimes y$ 是 $A\otimes B$ 对应于特征值 λu 的特征向量.

证 因为 x 是 A 对应于特征值 λ 的特征向量，y 是 B 对应于特征值 u 的特征向量，故

$$Ax = \lambda x, By = uy$$

因此

$$(A\otimes B)(x\otimes y) = (Ax)\otimes(By) = (\lambda x)\otimes(uy) = (\lambda u)(x\otimes y)$$

注意到 $x\otimes y\neq 0$，因此 $x\otimes y$ 是 $A\otimes B$ 对应于特征值 λu 的特征向量.

定理 5.3.5 设 $A\in C^{m\times m}$，$B\in C^{n\times n}$. 则存在 $m\times n$ 阶置换阵 P，使得

$$P^T(A\otimes B)P = B\otimes A$$

证 略.

定义 5.3.3 设 $A\in C^{m\times n}$，称

$$\underbrace{A\otimes A\otimes\cdots\otimes A}_{k个}$$

为 A 的关于 Kronecker 积的 k 次幂，记为 $A^{[k]}$.

定理 5.3.6 设 $A\in C^{m\times n}$，$B\in C^{n\times s}$. 则

$$(AB)^{[k]} = A^{[k]}B^{[k]}$$

利用数学归纳法容易验证.

定义 5.3.4 设 $A=(a_{ij})_{m\times n}\in C^{m\times n}$，称

$$\overrightarrow{A} = (a_{11}, a_{12}, \cdots, a_{1n}, a_{21}, a_{22}, \cdots, a_{2n}, \cdots, a_{m1}, a_{m2}, \cdots, a_{mn})^T$$

为 A 的（按行）拉直.

【例 5.3.2】 设 $A=\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix}$，则 $\overrightarrow{A}=(1, 2, 3, 4, 5, 6)^T$.

定理 5.3.7 设 $A\in C^{m\times n}$，$X\in C^{n\times p}$，$B\in C^{p\times s}$. 则

$$\overrightarrow{AXB} = (A \otimes B^{\mathrm{T}})\overrightarrow{X}$$

证 设$A=(a_{ij})_{m\times n}$，将X^{T}按列分块，即$X^{\mathrm{T}}=(x_1, x_2, \cdots, x_n)$，则有

$$AXB = \begin{pmatrix} (a_{11}x_1^{\mathrm{T}} + \cdots + a_{1n}x_n^{\mathrm{T}})B \\ \vdots \\ (a_{m1}x_1^{\mathrm{T}} + \cdots + a_{mn}x_n^{\mathrm{T}})B \end{pmatrix}, \overrightarrow{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

从而

$$\begin{aligned}\overrightarrow{AXB} &= ((a_{11}x_1^{\mathrm{T}} + \cdots + a_{1n}x_n^{\mathrm{T}})B, \cdots, (a_{11}x_1^{\mathrm{T}} + \cdots + a_{1n}x_n^{\mathrm{T}})B)^{\mathrm{T}} \\ &= \begin{pmatrix} B^{\mathrm{T}}(a_{11}x_1 + \cdots + a_{1n}x_n) \\ \vdots \\ B^{\mathrm{T}}(a_{11}x_1 + \cdots + a_{1n}x_n) \end{pmatrix} \\ &= \begin{pmatrix} a_{11}B^{\mathrm{T}} & \cdots & a_{1n}B^{\mathrm{T}} \\ \vdots & \vdots & \vdots \\ a_{m1}B^{\mathrm{T}} & \cdots & a_{mn}B^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\ &= (A \otimes B^{\mathrm{T}})\overrightarrow{X}\end{aligned}$$

推论 5.3.2 设$A\in \mathbf{C}^{m\times m}$，$B\in \mathbf{C}^{n\times n}$，$X\in \mathbf{C}^{m\times n}$，则：

(1) $\overrightarrow{AX} = (A \otimes E_n)\overrightarrow{X}, \overrightarrow{XB} = (E_m \otimes B^{\mathrm{T}})\overrightarrow{X}$；

(2) $\overrightarrow{AX + XB} = (A \otimes E_n + E_m \otimes B^{\mathrm{T}})\overrightarrow{X}$.

下面我们来讨论线性矩阵方程的求解问题.

在系统控制等工程领域中，经常遇到矩阵方程

$$AX + XB = F$$

的求解问题，其中$A\in \mathbf{C}^{m\times m}$，$B\in \mathbf{C}^{n\times n}$，$F\in \mathbf{C}^{m\times n}$，是已知矩阵.

由推论 5.3.2 (2)，有

$$(A \otimes E_n + E_m \otimes B^{\mathrm{T}})\overrightarrow{X} = \overrightarrow{F}$$

因此矩阵方程$AX+XB=F$有解的充分必要条件（$A\otimes E_n + E_m\otimes B^{\mathrm{T}}$）$\overrightarrow{X}=\overrightarrow{F}$有解. 这样，我们就将矩阵方程转化为通常的线性方程组的求解问题. 由线性代数的知识，可得如下结论.

定理 5.3.8 设$A\in \mathbf{C}^{m\times m}$，$B\in \mathbf{C}^{n\times n}$，$F\in \mathbf{C}^{m\times n}$. 矩阵方程$AX+XB=F$有解的充分必要条件是

$$\overrightarrow{F} \in R(A \otimes E_n + E_m \otimes B^{\mathrm{T}})$$

由定理 5.3.8、定理 5.3.4 (2)，可得如下定理.

定理 5.3.9 设$A\in \mathbf{C}^{m\times m}$，$B\in \mathbf{C}^{n\times n}$，$F\in \mathbf{C}^{m\times n}$. 矩阵方程$AX+XB=F$有唯一解的充分必要条件是$A$和$-B$没有相同的特征值.

【例 5.3.3】 解矩阵方程$AX+XB=F$，其中：

(1) $A=\begin{pmatrix} 1 & 2 \\ -2 & 5 \end{pmatrix}$，$B=\begin{pmatrix} 2 & 4 \\ 2 & 0 \end{pmatrix}$，$F=\begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix}$；

(2) $A=\begin{pmatrix}1 & -1\\ 0 & 2\end{pmatrix}$，$B=\begin{pmatrix}-3 & 4\\ 0 & -1\end{pmatrix}$，$F=\begin{pmatrix}0 & 5\\ 2 & -9\end{pmatrix}$.

解　(1) 由于 A 的特征值为 $\lambda_1=\lambda_2=3$，B 的特征值为 $u_1=2$，$u_2=-4$，由定理 5.3.9 知，矩阵方程 $AX+XB=F$ 有唯一解. 设 $X=\begin{bmatrix}x_1 & x_2\\ x_3 & x_4\end{bmatrix}$，可将矩阵方程 $AX+XB=F$ 转化为

$$\begin{bmatrix}3 & 2 & 2 & 0\\ 4 & 1 & 0 & 2\\ -2 & 0 & 7 & 2\\ 0 & -2 & 4 & 5\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\\ x_4\end{bmatrix}=\begin{bmatrix}1\\ 1\\ 2\\ 3\end{bmatrix}$$

解得 $x_1=-1$，$x_2=2$，$x_3=0$，$x_4=1$，因此矩阵方程 $AX+XB=F$ 的唯一解为

$$X=\begin{pmatrix}-1 & 2\\ 0 & 1\end{pmatrix}$$

(2) 由于 A 的特征值为 $\lambda_1=1$，$\lambda_2=2$，B 的特征值为 $u_1=-3$，$u_2=-1$. 设 $X=\begin{bmatrix}x_1 & x_2\\ x_3 & x_4\end{bmatrix}$，可将矩阵方程 $AX+XB=F$ 转化为

$$\begin{bmatrix}-2 & 0 & -1 & 0\\ 4 & 0 & 0 & -1\\ 0 & 0 & -1 & 0\\ 0 & 0 & 4 & 1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\\ x_4\end{bmatrix}=\begin{bmatrix}0\\ 5\\ 2\\ -9\end{bmatrix}$$

该方程组的通解是

$$\begin{bmatrix}x_1\\ x_2\\ x_3\\ x_4\end{bmatrix}=\begin{bmatrix}1\\ 0\\ -2\\ -1\end{bmatrix}+k\begin{bmatrix}0\\ 1\\ 0\\ 0\end{bmatrix}\text{，其中 } k \text{ 是任意常数.}$$

因此矩阵方程 $AX+XB=F$ 的通解为

$$X=\begin{pmatrix}1 & 0\\ -2 & -1\end{pmatrix}+k\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\text{，其中 } k \text{ 是任意常数.}$$

一般的线性矩阵方程为

$$\sum_{k=1}^{r}A_kXB_k=F$$

其中 $A_k\in C^{m\times n}$，$X\in C^{n\times p}$，$B_k\in C^{p\times s}$，$(k=1, 2, \cdots, r)$. 将上述方程拉直，利用拉直与直积的关系，可得它的等价线性方程组为

$$\left[\sum_{k=1}^{r}(A_k\otimes B_k^{T})\right]\vec{X}=\vec{F}$$

求解该方程组就可求得原方程的解.

§5.4　矩阵的微分和积分

在研究微分方程组时，为简化对问题的表达及求解过程，需要考虑以函数为元素的矩阵

的微分和积分. 在研究优化等问题时，则要遇到数量函数对向量变量或矩阵变量的导数，以及向量值或矩阵值函数对向量变量或矩阵变量的导数. 本节简单介绍这些内容.

定义 5.4.1 以变量 t 的函数为元素的矩阵 $\boldsymbol{A}(t)=(a_{ij}(t))_{m\times n}$ 称为函数矩阵. 若 $t\in[a,b]$，则称 $\boldsymbol{A}(t)$ 是定义在 $[a,b]$ 上的函数矩阵；若每个 $a_{ij}(t)$ 在 $[a,b]$ 上连续（可微、可积），则称 $\boldsymbol{A}(t)$ 在 $[a,b]$ 上连续（可微、可积）. 当 $\boldsymbol{A}(t)$ 在 $[a,b]$ 上可微时，规定其导数为

$$\boldsymbol{A}'(t)=(a_{ij}'(t))_{m\times n},\quad 或\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)=\left(\frac{d}{\mathrm{d}t}a_{ij}(t)\right)_{m\times n}$$

而当 $\boldsymbol{A}(t)$ 在 $[a,b]$ 上可积时，规定 $\boldsymbol{A}(t)$ 在 $[a,b]$ 上可积为

$$\int_a^b\boldsymbol{A}(t)\mathrm{d}t=\left(\int_a^b a_{ij}(t)\mathrm{d}t\right)_{m\times n}$$

【例 5.4.1】 设 $\boldsymbol{A}(t)=\begin{pmatrix}t^2+1 & \sin t & t\\ 0 & 1 & \cos t\end{pmatrix}$，求 $\boldsymbol{A}(t)$ 的导数和 $\int_0^1\boldsymbol{A}(t)\mathrm{d}t$.

解 $\boldsymbol{A}'(t)=\begin{pmatrix}2t & \cos t & 1\\ 0 & 0 & -\sin t\end{pmatrix}$,

$$\int_0^1\boldsymbol{A}(t)\mathrm{d}t=\begin{bmatrix}\int_0^1(t^2+1)\mathrm{d}t & \int_0^1\sin t\mathrm{d}t & \int_0^1 t\mathrm{d}t\\ 0 & \int_0^1 1\mathrm{d}t & \int_0^1\cos t\mathrm{d}t\end{bmatrix}$$

$$=\begin{bmatrix}\frac{4}{3} & 1-\cos 1 & \frac{1}{2}\\ 0 & 1 & \sin 1\end{bmatrix}$$

关于函数矩阵，类似于一元函数的求导法则，有如下求导法则.

定理 5.4.1 设 $\boldsymbol{A}(t)$、$\boldsymbol{B}(t)$ 是适当阶的可微矩阵，$u(t)$ 为可微函数，则：

(1) $[a\boldsymbol{A}(t)+b\boldsymbol{B}(t)]'=a\boldsymbol{A}'(t)+b\boldsymbol{B}'(t),(a,b\in\mathbf{C})$；

(2) $[u(t)\boldsymbol{A}(t)]'=u'(t)\boldsymbol{A}(t)+u(t)\boldsymbol{A}'(t)$；

(3) $[\boldsymbol{A}(t)\boldsymbol{B}(t)]'=\boldsymbol{A}'(t)\boldsymbol{B}(t)+\boldsymbol{A}(t)\boldsymbol{B}'(t)$；

(4) $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(u(t))=u'(t)\frac{\mathrm{d}}{\mathrm{d}u}\boldsymbol{A}(u)$；

(5) 当 $\boldsymbol{A}(t)$ 可逆时，$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}^{-1}(t)=-\boldsymbol{A}^{-1}(t)(\boldsymbol{A}'(t))\boldsymbol{A}^{-1}(t)$.

证 只证 (5)，其他可利用函数的求导法则直接验证. 由于 $\boldsymbol{A}(t)$ 可逆，因此 $\boldsymbol{A}(t)\boldsymbol{A}^{-1}(t)=\boldsymbol{E}$，两边对 t 求导，并利用 (3) 得

$$\boldsymbol{A}'(t)\boldsymbol{A}^{-1}(t)+\boldsymbol{A}(t)(\boldsymbol{A}^{-1}(t))'=0$$

从而

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}^{-1}(t)=-\boldsymbol{A}^{-1}(t)(\boldsymbol{A}'(t))\boldsymbol{A}^{-1}(t)$$

利用定义和一元函数积分的有关定理、性质，可得如下定理.

定理 5.4.2 设 $\boldsymbol{A}(t)$、$\boldsymbol{B}(t)$ 是 $[a,b]$ 上的适当阶的可积矩阵，$\boldsymbol{P}$，$\boldsymbol{Q}$ 是适当阶的常值矩阵，则：

(1) $\int_a^b[c_1\boldsymbol{A}(t)+c_2\boldsymbol{B}(t)]\mathrm{d}t=c_1\int_a^b\boldsymbol{A}(t)\mathrm{d}t+c_2\int_a^b\boldsymbol{B}(t)\mathrm{d}t,(c_1,c_2\in\mathbf{C})$；

(2) $\int_a^b \boldsymbol{P}\boldsymbol{A}(t)\mathrm{d}t = \boldsymbol{P}\left(\int_a^b \boldsymbol{A}(t)\mathrm{d}t\right), \int_a^b \boldsymbol{A}(t)\boldsymbol{Q}\mathrm{d}t = \left(\int_a^b \boldsymbol{A}(t)\mathrm{d}t\right)\boldsymbol{Q}$;

(3) 当 $\boldsymbol{A}(t)$ 在 $[a, b]$ 上连续时，对 $t \in (a,b)$，有 $\frac{\mathrm{d}}{\mathrm{d}t}\int_a^t \boldsymbol{A}(u)\mathrm{d}u = \boldsymbol{A}(t)$;

(4) 当 $\boldsymbol{A}(t)$ 在 $[a, b]$ 有连续导数时

$$\int_a^b \boldsymbol{A}'(t)\mathrm{d}t = \boldsymbol{A}(b) - \boldsymbol{A}(a)$$

习 题 5

5.1 构造一个 2×2 的可逆矩阵序列，其收敛于不可逆矩阵.

5.2 证明：当且仅当 $\boldsymbol{A}=\boldsymbol{E}$ 时，有 $\lim\limits_{n\to\infty}\boldsymbol{A}^n=\boldsymbol{E}$.

5.3 求 $\lim\limits_{n\to\infty}\begin{pmatrix}(1+n^{-1})^n & 0 \\ \dfrac{1+n}{1-n} & \sqrt[n]{n}\end{pmatrix}$.

5.4 求极限 $\lim\limits_{n\to\infty}\boldsymbol{A}^n$，其中 $\boldsymbol{A}=\begin{pmatrix}1/2 & 1 & 1 \\ 0 & 1/3 & 1 \\ 0 & 0 & 1/4\end{pmatrix}$.

5.5 如果 $\boldsymbol{B}=\lim\limits_{n\to\infty}\boldsymbol{A}^n$，则 $\boldsymbol{B}$ 是幂等阵，即 $\boldsymbol{B}^2=\boldsymbol{B}$.

5.6 设有一 5 阶复矩阵 $\boldsymbol{A}$ 的特征多项式为

$$(\lambda-1)^2\left(\lambda-\frac{1}{2}\right)^2\left(\lambda-\frac{1}{4}\right)$$

证明：$\lim\limits_{n\to\infty}\boldsymbol{A}^n$ 收敛的充分必要条件是 $\operatorname{rank}(\boldsymbol{A}-\boldsymbol{E})=3$.

5.7 设 $\boldsymbol{A}=\begin{pmatrix}0 & c & c \\ c & 0 & c \\ c & c & 0\end{pmatrix}$，讨论实数 c 取何值时，$\lim\limits_{n\to\infty}\boldsymbol{A}^n$ 为收敛矩阵.

5.8 设 $\boldsymbol{A}=\begin{pmatrix}4 & 2 & -5 \\ 1 & 4 & -9 \\ 1 & 3 & -7\end{pmatrix}$，$\boldsymbol{B}=\begin{pmatrix}3 & -1 \\ 1 & 1\end{pmatrix}$，求 $|\boldsymbol{A}\otimes\boldsymbol{B}|$，$\operatorname{tr}(\boldsymbol{A}\otimes\boldsymbol{B})$，以及 $\boldsymbol{A}\otimes\boldsymbol{B}$ 的特征值.

5.9 设 $\boldsymbol{A}\in\mathbf{R}^{m\times m}$ 的特征值为 λ_1，λ_2，…，λ_m，证明：

$$\boldsymbol{B}=\vec{1}\cdot\vec{1}^{\mathrm{T}}\otimes\boldsymbol{A}\quad(\vec{1}=(1,\cdots,1)^{\mathrm{T}}\in\mathbf{R}^n)$$

的特征值是 $n\lambda_1$，$n\lambda_2$，…，$n\lambda_m$ 和 m $(n-1)$ 重零.

5.10 解矩阵方程 $\boldsymbol{AX}+\boldsymbol{XB}=\boldsymbol{C}$，其中 $\boldsymbol{A}=\begin{pmatrix}2 & 6 & -15 \\ 1 & 1 & -5 \\ 1 & 2 & -6\end{pmatrix}$，$\boldsymbol{B}=\begin{pmatrix}3 & 1 \\ -4 & -1\end{pmatrix}$，$\boldsymbol{C}=\begin{pmatrix}1 & 4 \\ -2 & 6 \\ -1 & 4\end{pmatrix}$.

5.11 解矩阵方程 $\boldsymbol{AXB}=\boldsymbol{C}$，其中 $\boldsymbol{A}=\begin{pmatrix}2 & 6 & -15 \\ 1 & 1 & -5 \\ 1 & 5 & -10\end{pmatrix}$，$\boldsymbol{B}=\begin{pmatrix}3 & 1 \\ -4 & -1\end{pmatrix}$，$\boldsymbol{C}=\begin{pmatrix}-4 & 2 \\ 1 & -4 \\ -1 & 1\end{pmatrix}$.

5.12 证明：矩阵方程 $\boldsymbol{AX}+\boldsymbol{XB}=0$ 只有零解，其中 $\boldsymbol{A}=\begin{bmatrix}2 & 6 & 0\\1 & 1 & 0\\1 & 2 & -6\end{bmatrix}$，$\boldsymbol{B}=\begin{pmatrix}1 & 1\\-4 & -1\end{pmatrix}$.

5.13 若函数矩阵 $\boldsymbol{A}(x)=\begin{pmatrix}3x^2 & 1\\x-3 & 2x\end{pmatrix}$，求 $\|\boldsymbol{A}'(1)\|_2$.

5.14 设 $\boldsymbol{A}(x)=\begin{pmatrix}\sin x & \cos x-1\\x^2 & e^x-1\end{pmatrix}$，求 $\lim\limits_{x\to 0}\dfrac{\boldsymbol{A}(x)}{x}$，$\dfrac{\mathrm{d}\int_0^{x^2}\boldsymbol{A}(x)\mathrm{d}x}{\mathrm{d}x}$.

第六章 矩阵函数及其应用

§6.1 矩阵幂级数

在§5.2中，我们讨论了矩阵级数. 现在我们来讨论一类重要的矩阵级数——方阵的幂级数，因为它是建立矩阵函数的依据.

定义 6.1.1 设$\boldsymbol{A}\in\mathbf{C}^{n\times n}$，$\{c_k\}$是复数列，则称

$$\sum_{k=0}^{\infty}c_k\boldsymbol{A}^k$$

为方阵$\boldsymbol{A}$的幂级数.

定理 6.1.1 若幂级数$\sum\limits_{k=0}^{\infty}c_kz^k$的收敛半径是$r$. 则

(1) 若$\rho(\boldsymbol{A})<r$，则$\sum\limits_{k=0}^{\infty}c_k\boldsymbol{A}^k$绝对收敛；

(2) 若$\rho(\boldsymbol{A})>r$，则$\sum\limits_{k=0}^{\infty}c_k\boldsymbol{A}^k$发散.

其中$\rho(\boldsymbol{A})$是n阶矩阵$\boldsymbol{A}$的谱半径.

证 (1) 若$\rho(\boldsymbol{A})<r$，则存在某一矩阵范数$\|\cdot\|$，使得

$$\|\boldsymbol{A}\|\leqslant\frac{\rho(\boldsymbol{A})+r}{2}$$

因此

$$\|c_k\boldsymbol{A}^k\|\leqslant|c_k\|\cdot|\boldsymbol{A}\|^k\leqslant|c_k|\left(\frac{\rho(\boldsymbol{A})+r}{2}\right)^k$$

又因幂级数$\sum\limits_{k=0}^{\infty}c_kz^k$的收敛半径是$r$，因此$\sum\limits_{k=0}^{\infty}|c_k|\left(\frac{\rho(\boldsymbol{A})+r}{2}\right)^k$收敛，从而$\sum\limits_{k=0}^{\infty}\|c_k\boldsymbol{A}^k\|$收敛，故$\sum\limits_{k=0}^{\infty}c_k\boldsymbol{A}^k$绝对收敛.

(2) 设$\boldsymbol{A}$的特征值为λ_1，λ_2，…，λ_n，且$\rho(\boldsymbol{A})=|\lambda_1|$. 则存在非奇异阵$\boldsymbol{P}$，使得$\boldsymbol{A}=\boldsymbol{PJP}^{-1}$，其中$\boldsymbol{J}$是$\boldsymbol{A}$的Jordan标准形. $\boldsymbol{J}_1$是$\boldsymbol{A}$的以λ_1为特征值的Jordan块，由定理5.2.3知，$\sum\limits_{k=0}^{\infty}c_k\boldsymbol{A}^k$收敛的充分必要条件条件是$\sum\limits_{k=0}^{\infty}c_k\boldsymbol{J}^k$收敛. 若幂级数$\sum\limits_{k=0}^{\infty}c_kz^k$的收敛半径是$r$，则$\sum\limits_{k=0}^{\infty}c_k\lambda_1^k$发散，因此$\sum\limits_{k=0}^{\infty}c_k\boldsymbol{J}^k$发散，从而$\sum\limits_{k=0}^{\infty}c_k\boldsymbol{A}^k$发散.

推论 6.1.1 若幂级数$\sum\limits_{k=0}^{\infty}c_kz^k$在整个复平面上收敛，则对任意方阵$\boldsymbol{A}$，$\sum\limits_{k=0}^{\infty}c_k\boldsymbol{A}^k$绝对收敛.

推论 6.1.2 $\boldsymbol{E}+\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^k+\cdots$绝对收敛的充分必要条件是$\rho(\boldsymbol{A})<1$，且其和是$(\boldsymbol{E}-\boldsymbol{A})^{-1}$.

证 由于幂级数 $\sum\limits_{k=0}^{\infty} z^k$ 的收敛半径是 1，因此 $\rho(\boldsymbol{A})<1$，则 $\boldsymbol{E}+\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^k+\cdots$ 绝对收敛. 反之，若 $\boldsymbol{E}+\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^k+\cdots$ 绝对收敛，则 $\sum\limits_{k=1}^{\infty} \|\boldsymbol{A}^k\|$ 绝对收敛，故 $\|\boldsymbol{A}^k\|\to 0$，因此 $\rho(\boldsymbol{A})<1$.

因为

$$(\boldsymbol{E}+\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^k+\cdots)(\boldsymbol{E}-\boldsymbol{A})=\boldsymbol{E}$$

故

$$(\boldsymbol{E}-\boldsymbol{A})^{-1}=\boldsymbol{E}+\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^k+\cdots$$

定义 6.1.2 设 $\boldsymbol{A}=(a_{ij})_{n\times n}\in\mathbf{C}^{n\times n}$. 如果 $\sum\limits_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}|<|a_{ii}|$，$i=1, 2, \cdots, n$，则称 $\boldsymbol{A}$ 是行对角占优的. 如果 $\sum\limits_{\substack{j=1\\ j\neq i}}^{n} |a_{ji}|<|a_{ii}|$，$i=1, 2, \cdots, n$，则称 $\boldsymbol{A}$ 是列对角占优的.

定理 6.1.2 设 $\boldsymbol{A}=(a_{ij})_{n\times n}\in\mathbf{C}^{n\times n}$. 如果 $\boldsymbol{A}$ 是行（列）对角占优的，则 $\boldsymbol{A}$ 可逆.

证 由于

$$\boldsymbol{A}=\begin{pmatrix} a_{11} & 0 & \cdots & 0\\ 0 & a_{22} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & a_{nn}\end{pmatrix}\left(\begin{pmatrix} 1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{pmatrix}+\begin{pmatrix} 0 & \dfrac{a_{12}}{a_{11}} & \cdots & \dfrac{a_{1n}}{a_{11}}\\ \dfrac{a_{21}}{a_{22}} & 0 & \cdots & \dfrac{a_{2n}}{a_{22}}\\ \vdots & \vdots & \ddots & \vdots\\ \dfrac{a_{21}}{a_{nn}} & \dfrac{a_{n2}}{a_{nn}} & \cdots & 0\end{pmatrix}\right)$$

记

$$\boldsymbol{A}=\begin{pmatrix} a_{11} & 0 & \cdots & 0\\ 0 & a_{22} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & a_{nn}\end{pmatrix},\boldsymbol{B}=\begin{pmatrix} 0 & \dfrac{a_{12}}{a_{11}} & \cdots & \dfrac{a_{1n}}{a_{11}}\\ \dfrac{a_{21}}{a_{22}} & 0 & \cdots & \dfrac{a_{2n}}{a_{22}}\\ \vdots & \vdots & \ddots & \vdots\\ \dfrac{a_{21}}{a_{nn}} & \dfrac{a_{n2}}{a_{nn}} & \cdots & 0\end{pmatrix}$$

则 $\boldsymbol{A}$ 可逆. 由于 $\boldsymbol{A}$ 是行对角占优，因此 $\|\boldsymbol{B}\|_{\infty}<1$，故 $\boldsymbol{E}+\boldsymbol{B}$ 可逆，从而 $\boldsymbol{A}$ 可逆.

例如 $\boldsymbol{A}=\begin{pmatrix} 10 & -2 & 3 & -4\\ 4 & -8 & 1 & -1\\ -5 & 2 & 15 & 7\\ 5 & -1 & 0 & 9\end{pmatrix}$，由于 $\boldsymbol{A}$ 是行对角占优，因此 $\boldsymbol{A}$ 可逆.

§6.2 矩 阵 函 数

定义 6.2.1 设 $f(z)$ 的幂级数 $\sum\limits_{k=0}^{\infty} c_k z^k$ 的收敛半径是 r，则当矩阵 $\boldsymbol{A}$ 的谱半径 $\rho(\boldsymbol{A})<r$

时，把收敛的矩阵幂级数 $\sum_{k=0}^{\infty}c_kA^k$ 的和称为矩阵函数，记为 $f(A)$，即

$$f(A)=\sum_{k=0}^{\infty}c_kA^k$$

如

$$e^A=\sum_{k=0}^{\infty}\frac{1}{k!}A^k=E+A+\frac{1}{2!}A^2+\frac{1}{3!}A^3+\cdots$$

$$\cos A=E-\frac{1}{2!}A^2+\frac{1}{4!}A^4-\frac{1}{6!}A^6+\cdots$$

【例 6.2.1】 求微分方程组

$$\begin{cases}\dfrac{dx_1}{dt}=-2x_1+x_2+1\\ \dfrac{dx_2}{dt}=-4x_1+2x_2+2\\ \dfrac{dx_1}{dt}=x_1+x_3-1\end{cases}$$

满足初始条件 $x_1(0)=1$，$x_2(0)=1$，$x_3(0)=-1$ 的解.

解 微分方程组的满足初始条件 $x_1(0)=1$，$x_2(0)=1$，$x_3(0)=-1$ 的解为

$$x(t)=e^{At}\left(C_0+\int_0^t e^{-At}f(t)dt\right)$$

其中，$A=\begin{pmatrix}-2&1&0\\-4&2&0\\1&0&1\end{pmatrix}$，$C_0=(1,1,-1)^T$，$f(t)=(1,2,e^t-1)^T$（具体推导见本章第 4 节).

因为 $|\lambda E-A|=\lambda^3-\lambda^2$，由 Cayley-Hamilton 定理（定理 3.4.1）知，$A^3=A^2$，因此

$$A^{2+k}=A^2,k=2,3,\cdots$$

于是有

$$\begin{aligned}e^{At}&=\sum_{k=0}^{\infty}\frac{1}{k!}(At)^k=E+At+\frac{1}{2!}(At)^2+\frac{1}{3!}(At)^3+\frac{1}{4!}(At)^4+\cdots\\&=E+At+\left(\frac{1}{2!}t^2+\frac{1}{3!}t^3+\frac{1}{4!}t^4+\cdots\right)A^2=E+At+(e^t-1-t)A^2\\&=\begin{pmatrix}1-2t&t&0\\-4t&2t+1&0\\1+2t-e^t&e^t-t-1&e^t\end{pmatrix}\end{aligned}$$

因此

$$\begin{aligned}x(t)&=\begin{pmatrix}1-2t&t&0\\-4t&2t+1&0\\1+2t-e^t&e^t-t-1&e^t\end{pmatrix}\left(\begin{pmatrix}1\\1\\-1\end{pmatrix}+\int_0^t\begin{pmatrix}1+2t&-t&0\\4t&-2t+1&0\\1-2t-e^{-t}&e^{-t}+t-1&e^{-t}\end{pmatrix}\begin{pmatrix}1\\2\\e^t-1\end{pmatrix}\right)dt\\&=\begin{pmatrix}1-2t&t&0\\-4t&2t+1&0\\1+2t-e^t&e^t-t-1&e^t\end{pmatrix}\left(\begin{pmatrix}1\\1\\-1\end{pmatrix}+\begin{pmatrix}t\\2t\\0\end{pmatrix}\right)=\begin{pmatrix}1\\2\\(t-1)e^t\end{pmatrix}\end{aligned}$$

设 A 的 Jordan 标准形为 J. 由定理 5.2.3 知，若非奇异矩阵 P，使得 $A=PJP^{-1}$，则

$$f(\boldsymbol{A}) = \boldsymbol{P}f(\boldsymbol{J})\boldsymbol{P}^{-1}$$

如果

$$\boldsymbol{J} = \mathrm{diag}(\boldsymbol{J}_1, \boldsymbol{J}_2, \cdots, \boldsymbol{J}_s)$$

则有

$$\boldsymbol{J}^m = \mathrm{diag}(\boldsymbol{J}_1^m, \boldsymbol{J}_2^m, \cdots, \boldsymbol{J}_s^m)$$

因此

$$\begin{aligned} f(\boldsymbol{J}) &= \sum_{k=0}^{\infty} c_k \boldsymbol{J}^k = \mathrm{diag}\left(\sum_{k=0}^{\infty} c_k \boldsymbol{J}_1^m, \sum_{k=0}^{\infty} c_k \boldsymbol{J}_2^m, \cdots, \sum_{k=0}^{\infty} c_k \boldsymbol{J}_s^m\right) \\ &= \mathrm{diag}(f(\boldsymbol{J}_1), f(\boldsymbol{J}_2), \cdots, f(\boldsymbol{J}_s)) \end{aligned}$$

设

$$\boldsymbol{J}_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}_{k_i \times k_i}$$

令

$$\boldsymbol{H}_i = \begin{pmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}_{k_i \times k_i}$$

则

$$\boldsymbol{J}_i = \lambda_i \boldsymbol{E} + \boldsymbol{H}_i$$

而

$$\boldsymbol{H}_i^2 = \begin{pmatrix} 0 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & \ddots & 0 \\ & & & & 0 \end{pmatrix}$$

$$\cdots$$

$$\boldsymbol{H}_i^p = 0, \quad p \geqslant k_i$$

设函数 $f(z)$ 在 λ_i 处的 Taylor 展开式为

$$f(z) = \sum_{k=0}^{+\infty} \frac{f^{(k)}(\lambda_i)}{k!}(z - \lambda_i)^k$$

则有

$$\begin{aligned} f(\boldsymbol{J}_i) &= \sum_{k=0}^{+\infty} \frac{f^{(k)}(\lambda_i)}{k!}(\boldsymbol{J}_i - \lambda_i \boldsymbol{E})^k \\ &= \sum_{k=0}^{k_i} \frac{f^{(k)}(\lambda_i)}{k!}(\boldsymbol{J}_i - \lambda_i \boldsymbol{E})^k \\ &= f(\lambda_i)\boldsymbol{E} + \sum_{k=1}^{k_i} \frac{f^{(k)}(\lambda_i)}{k!} H_i^k \end{aligned}$$

故

$$f(\boldsymbol{J}_i)=\begin{pmatrix} f(\lambda_i) & \dfrac{f'(\lambda_i)}{1!} & \cdots & \dfrac{f^{(k_i-1)}(\lambda_i)}{(k_i-1)!} \\ & f(\lambda_i) & \ddots & \cdots \\ & & \ddots & \dfrac{f'(\lambda_i)}{1!} \\ & & & f(\lambda_i) \end{pmatrix}_{k_i\times k_i}$$

综上所述，可得利用 Jordan 标准形求矩阵函数的计算步骤如下：

(1) 求 $\boldsymbol{A}$ 的 Jordan 标准形为 $\boldsymbol{J}$ 及相应的相似变换阵 $\boldsymbol{P}$；

(2) 求 $f(\boldsymbol{J})=\text{diag}(f(\boldsymbol{J}_1),f(\boldsymbol{J}_2),\cdots,f(\boldsymbol{J}_s))$，其中

$$f(\boldsymbol{J}_i)=\begin{pmatrix} f(\lambda_i) & \dfrac{f'(\lambda_i)}{1!} & \cdots & \dfrac{f^{(k_i-1)}(\lambda_i)}{(k_i-1)!} \\ & f(\lambda_i) & \ddots & \cdots \\ & & \ddots & \dfrac{f'(\lambda_i)}{1!} \\ & & & f(\lambda_i) \end{pmatrix}_{k_i\times k_i}$$

(3) 计算 $f(\boldsymbol{A})=\boldsymbol{P}f(\boldsymbol{J})\boldsymbol{P}^{-1}$.

【例 6.2.2】 设 $\boldsymbol{A}=\begin{pmatrix}-4 & 2 & 10\\ -4 & 3 & 7\\ -3 & 1 & 7\end{pmatrix}$，求 $\cos\boldsymbol{A}$.

解 由例 3.3.10 可知，相似变换阵 $\boldsymbol{P}=\begin{pmatrix}2 & 0 & 1\\ 1 & 1 & -2\\ 1 & 0 & 1\end{pmatrix}$，使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}=\begin{pmatrix}2 & 1 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{pmatrix}$$

而

$$\cos\boldsymbol{J}=\begin{pmatrix}\cos 2 & -\sin 2 & -\dfrac{\cos 2}{2}\\ 0 & \cos 2 & -\sin 2\\ 0 & 0 & \cos 2\end{pmatrix}$$

因此

$$\begin{aligned}\cos\boldsymbol{A}&=\boldsymbol{P}(\cos\boldsymbol{J})\boldsymbol{P}^{-1}\\ &=\begin{pmatrix}2 & 0 & 1\\ 1 & 1 & -2\\ 1 & 0 & 1\end{pmatrix}\begin{pmatrix}\cos 2 & -\sin 2 & -\dfrac{\cos 2}{2}\\ 0 & \cos 2 & -\sin 2\\ 0 & 0 & \cos 2\end{pmatrix}\begin{pmatrix}2 & 0 & 1\\ 1 & 1 & -2\\ 1 & 0 & 1\end{pmatrix}^{-1}\\ &=\begin{pmatrix}2 & 0 & 1\\ 1 & 1 & -2\\ 1 & 0 & 1\end{pmatrix}\begin{pmatrix}\cos 2 & -\sin 2 & -\dfrac{\cos 2}{2}\\ 0 & \cos 2 & -\sin 2\\ 0 & 0 & \cos 2\end{pmatrix}\begin{pmatrix}1 & 0 & -1\\ -3 & 1 & 5\\ -1 & 0 & 2\end{pmatrix}\end{aligned}$$

$$
=\begin{pmatrix} 2\cos2+6\sin2 & -2\sin2 & -2\cos2-10\sin2 \\ 0.5\cos2+4\sin2 & \cos2-\sin2 & -\cos2-7\sin2 \\ 0.5\cos2+3\sin2 & -\sin2 & -5\sin2 \end{pmatrix}
$$

【例 6.2.3】 已知 $\boldsymbol{A}=\begin{pmatrix} 3 & 1 & -1 \\ -2 & 0 & 2 \\ -1 & -1 & 3 \end{pmatrix}$，求 $\sin\left(\frac{\pi}{4}\boldsymbol{A}\right)$，$e^{\boldsymbol{A}}$ 的 Jordan 标准形.

解 由于 $|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-2)^3$，而 $\operatorname{rank}(2\boldsymbol{E}-\boldsymbol{A})=1$，因此 $\boldsymbol{A}$ 的 Jordan 标准形为

$$
\boldsymbol{J}=\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}
$$

$$
\sin\left(\frac{\pi}{4}\boldsymbol{J}\right)=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},\quad e^{\boldsymbol{J}}=\begin{pmatrix} e^2 & 0 & 0 \\ 0 & e^2 & e^2 \\ 0 & 0 & e^2 \end{pmatrix}
$$

注意到存在非奇异阵 $\boldsymbol{P}$，使得 $f(\boldsymbol{A})=\boldsymbol{P}f(\boldsymbol{J})\boldsymbol{P}^{-1}$，即 $f(\boldsymbol{A})$ 与 $f(\boldsymbol{J})$ 相似，故 $\sin\left(\frac{\pi}{4}\boldsymbol{A}\right)$，$e^{\boldsymbol{A}}$ 的 Jordan 标准形分别是

$$
\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},\ \begin{pmatrix} e^2 & 0 & 0 \\ 0 & e^2 & 1 \\ 0 & 0 & e^2 \end{pmatrix}
$$

矩阵指数函数与三角函数是常用的矩阵函数，它们有些性质与普通的指数函数与三角函数相同. 但由于矩阵乘法不满足交换律，从而有些性质与一般的指数函数与三角函数不同.

定理 6.2.1 对任意 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，

(1) $\sin(-\boldsymbol{A})=-\sin\boldsymbol{A}, \cos(-\boldsymbol{A})=\cos\boldsymbol{A}$;

(2) $e^{i\boldsymbol{A}}=\cos\boldsymbol{A}+i\sin\boldsymbol{A}$;

(3) $\cos^2\boldsymbol{A}+\sin^2\boldsymbol{A}=\boldsymbol{E}$;

(4) $e^{\boldsymbol{A}+2\pi i\boldsymbol{E}}=e^{\boldsymbol{A}}$;

(5) $\frac{d}{dt}e^{\boldsymbol{A}t}=\boldsymbol{A}e^{\boldsymbol{A}t}=e^{\boldsymbol{A}t}\boldsymbol{A}$;

(6) $\frac{d}{dt}\sin(\boldsymbol{A}t)=\boldsymbol{A}\cos(\boldsymbol{A}t)=\cos(\boldsymbol{A}t)\boldsymbol{A}$;

(7) $\frac{d}{dt}\cos(\boldsymbol{A}t)=-\boldsymbol{A}\sin(\boldsymbol{A}t)=-\sin(\boldsymbol{A}t)\boldsymbol{A}$.

定理 6.2.2 设 $\boldsymbol{A}$，$\boldsymbol{B}\in\mathbf{C}^{n\times n}$. 若 $\boldsymbol{AB}=\boldsymbol{BA}$，则有

(1) $e^{\boldsymbol{A}+\boldsymbol{B}}=e^{\boldsymbol{B}+\boldsymbol{A}}=e^{\boldsymbol{A}}e^{\boldsymbol{B}}$;

(2) $\sin(\boldsymbol{A}+\boldsymbol{B})=\sin\boldsymbol{A}\cos\boldsymbol{B}+\cos\boldsymbol{A}\sin\boldsymbol{B}$;

(3) $\cos(\boldsymbol{A}+\boldsymbol{B})=\cos\boldsymbol{A}\cos\boldsymbol{B}-\sin\boldsymbol{A}\sin\boldsymbol{B}$.

证 这里只给出 (1) 的证明. 其他两个作为练习. 因为矩阵加法满足交换律，所以只需证明 $e^{\boldsymbol{A}+\boldsymbol{B}}=e^{\boldsymbol{A}}e^{\boldsymbol{B}}$. 由于 $\boldsymbol{AB}=\boldsymbol{BA}$，因此

$$
e^{\boldsymbol{A}}e^{\boldsymbol{B}}=\left(\sum_{k=0}^{+\infty}\frac{1}{k!}\boldsymbol{A}^k\right)\left(\sum_{k=0}^{+\infty}\frac{1}{k!}\boldsymbol{B}^k\right)
$$

$$= \boldsymbol{E} + (\boldsymbol{A}+\boldsymbol{B}) + \frac{1}{2!}(\boldsymbol{A}^2 + 2\boldsymbol{AB} + \boldsymbol{B}^2) + \cdots$$

$$= \sum_{k=0}^{+\infty} \frac{1}{k!}(\boldsymbol{A}+\boldsymbol{B})^k = \mathrm{e}^{\boldsymbol{A}+\boldsymbol{B}}$$

推论 6.2.1 $\sin(2\boldsymbol{A}) = 2\sin\boldsymbol{A}\cos\boldsymbol{A}, \cos(2\boldsymbol{A}) = \cos^2\boldsymbol{A} - \sin^2\boldsymbol{A}$.

若 $\boldsymbol{AB}\neq\boldsymbol{BA}$，则上述公式可能不成立. 如

$$\boldsymbol{A}=\begin{pmatrix}0&0\\1&0\end{pmatrix},\boldsymbol{B}=\begin{pmatrix}0&1\\0&0\end{pmatrix},\boldsymbol{AB}=\begin{pmatrix}0&0\\0&1\end{pmatrix}\neq\begin{pmatrix}1&0\\0&0\end{pmatrix}=\boldsymbol{BA};$$

$$\mathrm{e}^{\boldsymbol{A}}=\begin{pmatrix}1&0\\1&1\end{pmatrix},\mathrm{e}^{\boldsymbol{B}}=\begin{pmatrix}1&1\\0&1\end{pmatrix},\mathrm{e}^{\boldsymbol{A}+\boldsymbol{B}}=\begin{bmatrix}\mathrm{e}+\mathrm{e}^{-1}&\mathrm{e}-\mathrm{e}^{-1}\\\mathrm{e}-\mathrm{e}^{-1}&\mathrm{e}+\mathrm{e}^{-1}\end{bmatrix};$$

$$\mathrm{e}^{\boldsymbol{A}}\mathrm{e}^{\boldsymbol{B}}=\begin{pmatrix}1&1\\1&2\end{pmatrix}\neq\mathrm{e}^{\boldsymbol{B}}\mathrm{e}^{\boldsymbol{A}}=\begin{pmatrix}2&1\\1&1\end{pmatrix}$$

$$\mathrm{e}^{\boldsymbol{A}}\mathrm{e}^{\boldsymbol{B}}\neq\mathrm{e}^{\boldsymbol{A}+\boldsymbol{B}}\neq\mathrm{e}^{\boldsymbol{B}}\mathrm{e}^{\boldsymbol{A}}$$

定理 6.2.3 对任意 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，则有

(1) $\det(\mathrm{e}^{\boldsymbol{A}}) = \mathrm{e}^{\mathrm{tr}(\boldsymbol{A})}$;

(2) $(\mathrm{e}^{\boldsymbol{A}})^{-1} = \mathrm{e}^{-\boldsymbol{A}}$.

证 (1) 设 $\boldsymbol{J}$ 是 $\boldsymbol{A}$ 的 Jordan 标准形，λ_1，λ_2，…，λ_n 为 $\boldsymbol{A}$ 的特征值，则存在非奇异阵 $\boldsymbol{P}$，使得 $\boldsymbol{A}=\boldsymbol{PJP}^{-1}$，因此

$$\mathrm{e}^{\boldsymbol{A}} = \boldsymbol{P}\mathrm{e}^{\boldsymbol{J}}\boldsymbol{P}^{-1}$$

$$\det(\mathrm{e}^{\boldsymbol{A}}) = \det(\mathrm{e}^{\boldsymbol{J}}) = \mathrm{e}^{\lambda_1+\lambda_2+\cdots+\lambda_n} = \mathrm{e}^{\mathrm{tr}(\boldsymbol{A})}$$

(2) 由于

$$\mathrm{e}^{\boldsymbol{A}}\mathrm{e}^{-\boldsymbol{A}} = \mathrm{e}^{\boldsymbol{A}-\boldsymbol{A}} = \mathrm{e}^{\boldsymbol{0}} = \boldsymbol{E}$$

因此

$$(\mathrm{e}^{\boldsymbol{A}})^{-1} = \mathrm{e}^{-\boldsymbol{A}}$$

§ 6.3 矩阵函数的一般定义及其计算

利用定义 6.2.1 定义的矩阵函数，其实质就是先将数量函数 $f(z)$ 展开成收敛的幂级数，然后以矩阵 $\boldsymbol{A}$ 代替 z，得到 $f(\boldsymbol{A})$. 但是，对于任意给定的函数 $f(z)$ 要求能展开成收敛的幂级数条件太强，一般不易满足. 从利用 Jordan 标准形求矩阵函数的方法可知，矩阵函数 $f(\boldsymbol{A})$ 只与函数 $f(z)$ 以及它的一定阶导数在 $\boldsymbol{A}$ 的特征值的取值有关. 因此我们可以给出矩阵函数的一种较为广泛的定义. 先给出一些概念.

定义 6.3.1 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$的最小多项式为 $\varphi(\lambda) = (\lambda-\lambda_1)^{m_1}\cdots(\lambda-\lambda_s)^{m_s}$，则称集合

$$\{(\lambda_i,m_i) \mid i=1,2,\cdots,s\}$$

为 $\boldsymbol{A}$ 的谱. 记为 $\sigma_{\boldsymbol{A}}$.

定义 6.3.2 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$的最小多项式为 $\varphi(\lambda) = (\lambda-\lambda_1)^{m_1}\cdots(\lambda-\lambda_s)^{m_s}$，称

$$\{f^{(k_i)}(\lambda_i) \mid k_i = 0,1,\cdots,m_i-1; i=1,2,\cdots,s\}$$

为函数 $f(z)$ 在 $\boldsymbol{A}$ 上的谱值. 记为 $f(\sigma_{\boldsymbol{A}})$.

例如，如果 $\boldsymbol{A}$ 的最小多项式为 $\varphi(\lambda) = (\lambda-3)(\lambda-2)^3$，则 $\boldsymbol{A}$ 的谱是 $\{(3,1),(2,3)\}$，

函数 $\sin z$ 在 $\boldsymbol{A}$ 上的谱值是

$$\sin 3, \sin 2, \cos 2, -\sin 2$$

定义 6.3.3 对任意函数 $f(z)$，若存在复系数多项式 $g(z)$，使得

$$f(\sigma_{\boldsymbol{A}}) = g(\sigma_{\boldsymbol{A}})$$

则定义矩阵函数 $f(\boldsymbol{A})$ 为 $g(\boldsymbol{A})$，即

$$f(\boldsymbol{A}) = g(\boldsymbol{A})$$

下面我们来介绍 $g(z)$ 的确定方法. 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$ 的最小多项式为 $\varphi(\lambda)=(\lambda-\lambda_1)^{m_1}\cdots(\lambda-\lambda_s)^{m_s}$，复系数多项式 $g(z)$，使得 $f(\sigma_{\boldsymbol{A}})=g(\sigma_{\boldsymbol{A}})$，则存在复系数多项式 $h(z)$，$r(z)$，使得

$$g(z) = \varphi(z)h(z) + r(z)$$

其中 $r(z) = a_0 + a_1 z + \cdots + a_{m-1}z^{m-1}$，$m=m_1+m_2+\cdots+m_s$，因此

$$f(\boldsymbol{A}) = g(\boldsymbol{A}) = \varphi(\boldsymbol{A})h(\boldsymbol{A}) + r(\boldsymbol{A}) = r(\boldsymbol{A})$$

另一方面，利用 $f(\sigma_{\boldsymbol{A}}) = g(\sigma_{\boldsymbol{A}})$ 可得 m 个方程，因此可唯一确定 m 个系数. 这样一来，可设

$$g(z) = r(z) = a_0 + a_1 z + \cdots + a_{m-1}z^{m-1}$$

由此，可得求矩阵函数的待定系数法：

(1) 求 $\boldsymbol{A}$ 的最小多项式为 $\varphi(\lambda) = (\lambda-\lambda_1)^{m_1}\cdots(\lambda-\lambda_s)^{m_s}$；

(2) 设 $r(z) = a_0 + a_1 z + \cdots + a_{m-1}z^{m-1}$，其中，$m = \sum_{k=1}^{s} m_k$. 利用 $f(\sigma_{\boldsymbol{A}}) = r(\sigma_{\boldsymbol{A}})$ 得到 m 个方程，求解该方程组，确定 $r(z)$ 的系数 a_0，a_1，…，a_{m-1}；

(3) $f(\boldsymbol{A}) = r(\boldsymbol{A})$.

【例 6.3.1】 设 $\boldsymbol{A}=\begin{pmatrix}-4 & 2 & 10\\ -4 & 3 & 7\\ -3 & 1 & 7\end{pmatrix}$ 求 $\cos\boldsymbol{A}$.

解 由于 $|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-2)^3$，而 $\operatorname{rank}(2\boldsymbol{E}-\boldsymbol{A})=2$，因此 $\boldsymbol{A}$ 的最小多项式

$$\varphi(\lambda) = (\lambda-2)^3$$

令

$$r(z) = a_0 + a_1 z + a_2 z^2$$

由 $\cos(\sigma_{\boldsymbol{A}}) = r(\sigma_{\boldsymbol{A}})$，可得

$$\begin{cases} a_0 + 2a_1 + 4a_2 = \cos 2 \\ a_1 + 4a_2 = -\sin 2 \\ 2a_2 = -\cos 2 \end{cases}$$

解得 $a_0=-\cos 2+2\sin 2$，$a_1=-\sin 2+2\cos 2$，$a_2=-\dfrac{\cos 2}{2}$. 因此

$$\cos\boldsymbol{A} = (-\cos 2 + 2\sin 2)\boldsymbol{E} + (-\sin 2 + 2\cos 2)\boldsymbol{A} - \frac{\cos 2}{2}\boldsymbol{A}^2$$

$$= \begin{pmatrix} 2\cos 2 + 6\sin 2 & -2\sin 2 & -2\cos 2 - 10\sin 2 \\ 0.5\cos 2 + 4\sin 2 & \cos 2 - \sin 2 & -\cos 2 - 7\sin 2 \\ 0.5\cos 2 + 3\sin 2 & -\sin 2 & -5\sin 2 \end{pmatrix}$$

【例 6.3.2】 设 $\boldsymbol{A}=\begin{pmatrix}3 & -3 & 2\\ -1 & 5 & -2\\ -1 & 3 & 0\end{pmatrix}$，求 $\ln\boldsymbol{A}$.

解　因为 A 的特征多项式为

$$|\lambda E - A| = (\lambda - 2)^2(\lambda - 4)$$

因此 A 的最小多项式 $\varphi(\lambda)$ 只可能是

$$\varphi(\lambda) = (\lambda - 2)(\lambda - 4) \quad 或 \quad \varphi(\lambda) = (\lambda - 2)^2(\lambda - 4)$$

由于 $(A - 2E)(A - 4E) = 0$，因此 $\varphi(\lambda) = (\lambda - 2)(\lambda - 4)$.

令

$$r(z) = a_0 + a_1 z$$

由

$$\begin{cases} a_0 + 2a_1 = \ln 2 \\ a_0 + 3a_1 = \ln 3 \end{cases}$$

可得 $a_1 = \ln 3 - \ln 2$，$a_0 = 3\ln 2 - 2\ln 3$，故

$$\ln A = (3\ln 2 - 2\ln 3)E + (\ln 3 - \ln 2)\begin{pmatrix} 3 & -3 & 2 \\ -1 & 5 & -2 \\ -1 & 3 & 0 \end{pmatrix}$$

由求矩阵函数的待定系数法

$$f(A) = r(A) = a_0 + a_1 A + \cdots + a_{m-1} A^{m-1}$$

利用 $f(\sigma_A) = r(\sigma_A)$ 得到 m 个方程，再利用克莱姆法则求解该方程组，可以确定 $r(z)$ 的系数 a_0，a_1，…，a_{m-1}，即

$$a_i = \sum_{i=1}^{s} \sum_{j=0}^{m_j - 1} c_{ij} f^{(j)}(\lambda_i)$$

其中 c_{ij} 仅与 A 有关，而与函数 f 无关. 因此

$$f(A) = \sum_{i=1}^{s} \sum_{j=0}^{m_j - 1} f^{(j)}(\lambda_i) A_{ij}$$

其中 A_{ij} 仅与 A 有关，而与函数 f 无关. 这样，如果利用一些特殊的 f 将 A_{ij} 求出，就可求出要求的 $f(A)$. 从而得到如下求矩阵函数的待定矩阵法的计算步骤.

(1) 求 A 的最小多项式为 $\varphi(\lambda) = (\lambda - \lambda_1)^{m_1} \cdots (\lambda - \lambda_s)^{m_s}$；

(2) 令 $f(A) = \sum_{i=1}^{s} \sum_{j=0}^{m_j - 1} f^{(j)}(\lambda_i) A_{ij}$，取一些特殊的 f 求出 A_{ij}；

(3) 求要计算的 $f(A)$.

【例 6.3.3】　已知 $A = \begin{pmatrix} 3 & 1 & -1 \\ -2 & 0 & 2 \\ -1 & -1 & 3 \end{pmatrix}$，求 $\sin\left(\frac{\pi}{4}A\right)$ 和 e^A.

解　由于 $|\lambda E - A| = (\lambda - 2)^3$，而 $\mathrm{rank}(2E - A) = 1$，因此 A 的最小多项式

$$\varphi(\lambda) = (\lambda - 2)^2$$

令

$$f(A) = f(2)A_1 + f'(2)A_2$$

取 $f(z) = 1$，则有 $A_1 = E$；取 $f(z) = z - 2$，则有 $A_2 = A - 2E$. 因此

$$f(A) = f(2)E + f'(2)(A - 2E)$$

特别地

$$\sin\left(\frac{\pi}{4}A\right)=E$$

$$e^{A}=e^{2}E+e^{2}(A-2E)=e^{2}(A-E)$$

$$=e^{2}\begin{bmatrix}2 & 1 & -1\\ -2 & -1 & 2\\ -1 & -1 & 2\end{bmatrix}$$

【例 6.3.4】 已知 $A=\begin{bmatrix}2 & 2 & 1\\ 1 & 3 & 1\\ 1 & 2 & 2\end{bmatrix}$，求 e^{3A}.

解 因为 $|\lambda E-A|=(\lambda-1)^2(\lambda-5)$，$\text{rank}(A-E)=1$，因此 A 的最小多项式

$$\varphi(\lambda)=(\lambda-1)(\lambda-5)$$

(1) Jordan 标准形法. 由于 A 的最小多项式无重根，因此 A 的 Jordan 标准形是对角阵，因此 A 的 Jordan 标准形为

$$J=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 5\end{bmatrix}$$

经计算，可得相应的相似变换阵 P 为

$$\begin{bmatrix}1 & 2 & 1\\ 0 & -1 & 1\\ -1 & 0 & 1\end{bmatrix}$$

因此

$$e^{3A}=\begin{bmatrix}1 & 2 & 1\\ 0 & -1 & 1\\ -1 & 0 & 1\end{bmatrix}\begin{bmatrix}e^3 & 0 & 0\\ 0 & e^3 & 0\\ 0 & 0 & e^{15}\end{bmatrix}\begin{bmatrix}1 & 2 & 1\\ 0 & -1 & 1\\ -1 & 0 & 1\end{bmatrix}^{-1}$$

$$=\begin{bmatrix}1 & 2 & 1\\ 0 & -1 & 1\\ -1 & 0 & 1\end{bmatrix}\begin{bmatrix}e^3 & 0 & 0\\ 0 & e^3 & 0\\ 0 & 0 & e^{15}\end{bmatrix}\begin{bmatrix}\frac{1}{4} & \frac{1}{2} & \frac{-3}{4}\\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{4}\\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4}\end{bmatrix}$$

$$=\begin{bmatrix}\frac{3}{4}e^3+\frac{1}{4}e^{15} & -\frac{1}{2}e^3+\frac{1}{2}e^{15} & -\frac{1}{4}e^3+\frac{1}{4}e^{15}\\ -\frac{1}{4}e^3+\frac{1}{4}e^{15} & \frac{1}{2}e^3+\frac{1}{2}e^{15} & -\frac{1}{4}e^3+\frac{1}{4}e^{15}\\ -\frac{1}{4}e^3+\frac{1}{4}e^{15} & -\frac{1}{2}e^3+\frac{1}{2}e^{15} & \frac{3}{4}e^3+\frac{1}{4}e^{15}\end{bmatrix}$$

(2) 待定系数法. 由于 A 的最小多项式 $\varphi(\lambda)=(\lambda-1)(\lambda-5)$，可令 $r(z)=a_0+a_1z$，由

$$\begin{cases}a_0+a_1=e^3\\ a_0+5a_1=e^{15}\end{cases}$$

解得：$a_0=\frac{1}{4}(5e^3-e^{15})$，$a_1=\frac{1}{4}(e^{15}-e^3)$. 因此

$$e^{3A}=\frac{1}{4}(5e^3-e^{15})E+\frac{1}{4}(e^{15}-e^3)A$$

(3) 待定矩阵法．由于 A 的最小多项式 $\varphi(\lambda)=(\lambda-1)(\lambda-5)$，可令

$$f(A)=f(1)A_1+f(5)A_2$$

取 $f(z)=z-1$，则有 $A_2=\frac{1}{4}(A-E)$；取 $f(z)=z-5$，则有 $A_1=-\frac{1}{4}(A-5E)$．

故

$$\begin{aligned}f(A)&=-\frac{1}{4}f(1)(A-5E)+\frac{1}{4}f(5)(A-E)\\&=\frac{1}{4}[f(5)-f(1)]A+\frac{1}{4}[5f(1)-f(5)]E.\end{aligned}$$

特别地

$$e^{3A}=\frac{1}{4}[e^{15}-e^3]A+\frac{1}{4}[5e^3-e^{15}]E.$$

在 Matlab 中，对常用的函数，可以直接计算矩阵函数．

【例 6.3.5】 设 $A=\begin{pmatrix}-1&2&2\\3&-1&1\\2&2&-1\end{pmatrix}$，求 $\sin A$ 和 e^A．

解

```
>>sin(A)
ans=
   -0.8415    0.9093    0.9093
    0.1411   -0.8415    0.8415
    0.9093    0.9093   -0.8415
>>exp(A)
ans=
    0.3679    7.3891    7.3891
   20.0855    0.3679    2.7183
    7.3891    7.3891    0.3679
```

§6.4 矩阵函数的应用

本节中，我们介绍矩阵函数及矩阵微积分的一些应用．我们先讨论一阶线性常系数微分方程组的初值问题的求解．

在数学或工程技术中，经常要研究一阶常系数线性微分方程组

$$\begin{cases}\dfrac{dx_1(t)}{dt}=a_{11}x_1(t)+a_{12}x_2(t)+\cdots+a_{1n}x_n(t)+f_1(t)\\\dfrac{dx_2(t)}{dt}=a_{21}x_1(t)+a_{22}x_2(t)+\cdots+a_{2n}x_n(t)+f_2(t)\\\cdots\\\dfrac{dx_n(t)}{dt}=a_{n1}x_1(t)+a_{n2}x_2(t)+\cdots+a_{nn}x_n(t)+f_n(t)\end{cases}\tag{6.4.1}$$

满足初始条件

$$x_i(t_0)=c_i,\quad i=1,2,\cdots,n \tag{6.4.2}$$

的解.

如果记 $\boldsymbol{A}=(a_{ij})_{n\times n}$，$\boldsymbol{c}=(c_1,\ c_2,\ \cdots,\ c_n)^{\mathrm{T}}$，

$$\boldsymbol{x}(t)=(x_1(t),x_2(t),\cdots,x_n(t))^{\mathrm{T}},\boldsymbol{f}(t)=(f_1(t),f_2(t),\cdots,f_n(t))^{\mathrm{T}}$$

则上述问题可写成

$$\begin{cases}\dfrac{\mathrm{d}x(t)}{\mathrm{d}t}=\boldsymbol{Ax}(t)+\boldsymbol{f}(t)\\ \boldsymbol{x}(t_0)=\boldsymbol{c}\end{cases} \tag{6.4.3}$$

由于 $\boldsymbol{A}$ 是常值矩阵，因此

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{x}(t))=\mathrm{e}^{-\boldsymbol{A}t}(-\boldsymbol{A})\boldsymbol{x}(t)+\mathrm{e}^{-\boldsymbol{A}t}\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}(t)$$

$$=\mathrm{e}^{-\boldsymbol{A}t}\left[\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}-\boldsymbol{Ax}(t)\right]=\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{f}(t)$$

将上式两边在 $[t_0,\ t]$ 上积分，得到

$$\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{x}(t)-\mathrm{e}^{-\boldsymbol{A}t_0}\boldsymbol{x}(t_0)=\int_0^t\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{f}(t)\mathrm{d}t$$

因此微分方程组的初值问题的解为

$$x(t)=\mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{c}+\mathrm{e}^{\boldsymbol{A}t}\int_0^t\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{f}(t)\mathrm{d}t \tag{6.4.4}$$

【例 6.4.1】 求如下初值问题的解

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\boldsymbol{A}x\\ \boldsymbol{x}(0)=(1,1,1)^{\mathrm{T}}\end{cases},\ \boldsymbol{A}=\begin{pmatrix}3&-1&1\\2&0&-1\\1&-1&2\end{pmatrix}$$

解 $$|\lambda\boldsymbol{E}-\boldsymbol{A}|=\begin{vmatrix}\lambda-3&1&-1\\-2&\lambda&1\\-1&1&\lambda-2\end{vmatrix}=\lambda(\lambda-2)(\lambda-3)$$

故 $\boldsymbol{A}$ 有三个不同的特征值，从而 $\boldsymbol{A}$ 可与对角形矩阵相似．与三个特征值 0，2，3 相应的三个线性无关的特征向量为

$$\boldsymbol{p}_1=(1,5,2)^{\mathrm{T}},\boldsymbol{p}_2=(1,1,0)^{\mathrm{T}},\boldsymbol{p}_3=(2,1,1)^{\mathrm{T}}$$

从而

$$\boldsymbol{P}=\begin{pmatrix}1&1&2\\5&1&1\\2&0&1\end{pmatrix},\boldsymbol{P}^{-1}=-\frac{1}{6}\begin{pmatrix}1&-1&-1\\-3&-3&9\\-2&2&4\end{pmatrix}$$

使得 $\boldsymbol{P}^{-1}\boldsymbol{AP}=\begin{pmatrix}0&0&0\\0&2&0\\0&0&3\end{pmatrix}$.

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}(0)=\boldsymbol{P}\begin{pmatrix}1&0&0\\0&\mathrm{e}^{2t}&0\\0&0&\mathrm{e}^{3t}\end{pmatrix}\boldsymbol{P}^{-1}\boldsymbol{x}(0)=-\frac{1}{6}\begin{pmatrix}-1+3\mathrm{e}^{2t}-8\mathrm{e}^{3t}\\-5+3\mathrm{e}^{2t}-4\mathrm{e}^{3t}\\-2-4\mathrm{e}^{32}\end{pmatrix}$$

【例 6.4.2】　设 $A=\begin{pmatrix}2&0&0\\1&1&1\\1&-1&3\end{pmatrix}$，求如下初值问题的解.

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}=\boldsymbol{A}(t)\boldsymbol{x}(t)+(1,0,-1)^{\mathrm{T}}\\ \boldsymbol{x}(0)=(1,1,1)^{\mathrm{T}}\end{cases}$$

解　因为

$$\mathrm{e}^{At}=\mathrm{e}^{2t}\begin{pmatrix}1&0&0\\t&1-t&t\\t&-t&1+t\end{pmatrix}$$

所以

$$\int_0^t \mathrm{e}^{-At}\boldsymbol{f}(t)\mathrm{d}t=\int_0^t \mathrm{e}^{-2t}\begin{pmatrix}1\\0\\-1\end{pmatrix}\mathrm{d}t=\frac{1}{2}\begin{pmatrix}1-\mathrm{e}^{-2t}\\0\\\mathrm{e}^{-2t}-1\end{pmatrix}$$

$$\begin{aligned}\boldsymbol{x}(t)&=\mathrm{e}^{A(t-t_0)}\boldsymbol{c}+\mathrm{e}^{At}\int_{t_0}^t \mathrm{e}^{-At}\boldsymbol{f}(t)\mathrm{d}t\\&=\mathrm{e}^{2t}\begin{pmatrix}1&0&0\\t&1-t&t\\t&-t&1+t\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix}+\frac{1}{2}\mathrm{e}^{2t}\begin{pmatrix}1&0&0\\t&1-t&t\\t&-t&1+t\end{pmatrix}\begin{pmatrix}1-\mathrm{e}^{-2t}\\0\\\mathrm{e}^{-2t}-1\end{pmatrix}\\&=\frac{\mathrm{e}^{2t}}{2}\begin{pmatrix}3-\mathrm{e}^{-2t}\\2t+2\\2t+1+\mathrm{e}^{-2t}\end{pmatrix}\end{aligned}$$

【例 6.4.3】　设 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 为常数矩阵，$\boldsymbol{X}=(x_{ij}(t))_{n\times n}$，$a$ 为常数．求 Cauchy 微分方程组

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t}=\frac{1}{t-a}\boldsymbol{AX}$$

的通解.

解　作变量代换 $u=\ln(t-a)$，即 $t=\mathrm{e}^u+a$，则

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}u}=\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}u}=\frac{1}{t-a}\boldsymbol{AX}(t-a)=\boldsymbol{AX}$$

$$\boldsymbol{X}(t)=\mathrm{e}^{Au}\boldsymbol{C}=\mathrm{e}^{A\ln(t-a)}\boldsymbol{C}=(t-a)^{A}\boldsymbol{C}$$

其中 $\boldsymbol{C}$ 为 n 阶常数矩阵.

设 a_1，a_2，…，a_n 是常数，$f(t)$ 为已知函数，称

$$y^{(n)}+a_1y^{(n-1)}+a_2y^{(n-2)}+\cdots+a_{n-1}y'+a_ny=u(t) \tag{6.4.5}$$

为 n 阶常系数微分方程.

由于常系数线性微分方程组的矩阵形式解已经得到，因此，我们可以利用如下方式将 n 阶常系数微分方程转化为常系数线性微分方程组，进而求出它的解.

令

$$x_i(t)=y^{(i-1)}(t),\quad i=1,2,\cdots,n$$

则有

$$\begin{cases} x'_1 = x_2 \\ x'_2 = x_3 \\ \vdots \\ x'_{n-1} = x_n \\ x'_n = -a_n x_1 - a_{n-1} x_2 - \cdots - a_1 x_n + u(t) \end{cases} \tag{6.4.6}$$

若记

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{pmatrix}$$

$$\boldsymbol{x}(t) = (x_1(t), x_2(t), \cdots, x_n(t))^{\mathrm{T}}, \boldsymbol{f}(t) = (0, 0, \cdots, 0, u(t))^{\mathrm{T}}, \boldsymbol{c} = (c_1, c_2, \cdots, c_n)^{\mathrm{T}}$$

则有

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{f}(t)$$

因此 n 阶常系数微分方程的定解问题

$$\begin{cases} y^{(n)} + a_1 y^{(n-1)} + a_2 y^{(n-2)} + \cdots + a_{n-1} y' + a_n y = u(t) \\ y^{(i)}(0) = c_i, i = 0, 1, \cdots, n-1 \end{cases} \tag{6.4.7}$$

可转化为如下常系数线性微分方程组的初值问题：

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{A}(t)\boldsymbol{x}(t) + \boldsymbol{f}(t) \\ \boldsymbol{x}(t_0) = \boldsymbol{c} \end{cases} \tag{6.4.8}$$

它的解为

$$\boldsymbol{x}(t) = \mathrm{e}^{\boldsymbol{A}t}\boldsymbol{c} + \mathrm{e}^{\boldsymbol{A}t}\int_0^t \mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{f}(t)\mathrm{d}t$$

由于定解问题（6.4.7）的解是初值问题

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{A}(t)\boldsymbol{x}(t) + \boldsymbol{f}(t) \\ \boldsymbol{x}(0) = \boldsymbol{c} \end{cases}$$

的解的第一个分量，因此定解问题的解为

$$y(t) = (1, 0, 0, \cdots, 0)(\mathrm{e}^{\boldsymbol{A}t}c + \mathrm{e}^{\boldsymbol{A}t}\int_0^t \mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{f}(t)\mathrm{d}t) \tag{6.4.9}$$

【例 6.4.4】 求如下常系数微分方程的定解问题

$$\begin{cases} y''' + 7y'' + 14y' + 8y = 1 \\ y''(0) = y'(0) = y(0) = 0 \end{cases}$$

的解.

解 令

$$\begin{cases} x_1 = y \\ x_2 = y', \\ x_3 = y'' \end{cases} \boldsymbol{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -8 & -14 & -7 \end{pmatrix}, \boldsymbol{f} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

则定解问题

$$\begin{cases} y''' + 7y'' + 14y' + 8y = 1 \\ y''(0) = y'(0) = y(0) = 0 \end{cases}$$

可转化为如下常系数线性微分方程组的初值问题

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{A}(t)\boldsymbol{x}(t) + \boldsymbol{f}(t) \\ \boldsymbol{x}(0) = (0,0,0)^{\mathrm{T}} \end{cases}$$

由于

$$|\lambda\boldsymbol{E} - \boldsymbol{A}| = (\lambda+1)(\lambda+2)(\lambda+4)$$

因此 $\boldsymbol{A}t$ 的特征值为 $\lambda_1=-t$，$\lambda_2=-2t$，$\lambda_3=-4t$，故

$$g(t\boldsymbol{A}) = g(-t)\boldsymbol{A}_1 + g(-2t)\boldsymbol{A}_2 + g(-4t)\boldsymbol{A}_3$$

取 $g(z)=z+t$，则有 $t\boldsymbol{A}+t\boldsymbol{E}=-t\boldsymbol{A}_2-3t\boldsymbol{A}_3$；取 $g(z)=z+2t$，则有 $t\boldsymbol{A}+2t\boldsymbol{E}=t\boldsymbol{A}_1-2t\boldsymbol{A}_3$；取 $g(z)=(z+t)(z+2t)$，则有 $(t\boldsymbol{A}+t\boldsymbol{E})(t\boldsymbol{A}+2t\boldsymbol{E})=6t^2\boldsymbol{A}_3$.

由它们可解得

$$\boldsymbol{A}_1 = \frac{1}{3}\begin{pmatrix} 8 & 6 & 1 \\ -8 & -6 & -1 \\ 8 & 6 & 1 \end{pmatrix}, \boldsymbol{A}_2 = \frac{1}{2}\begin{pmatrix} -4 & -5 & -1 \\ 8 & 10 & 2 \\ -16 & -20 & 4 \end{pmatrix}, \boldsymbol{A}_3 = \frac{1}{6}\begin{pmatrix} 2 & 3 & 1 \\ -8 & -12 & -4 \\ 32 & 48 & 16 \end{pmatrix}$$

$$\mathrm{e}^{t\boldsymbol{A}} = \frac{\mathrm{e}^{-t}}{3}\begin{pmatrix} 8 & 6 & 1 \\ -8 & -6 & -1 \\ 8 & 6 & 1 \end{pmatrix} + \frac{\mathrm{e}^{-2t}}{2}\begin{pmatrix} -4 & -5 & -1 \\ 8 & 10 & 2 \\ -16 & -20 & 4 \end{pmatrix} + \frac{\mathrm{e}^{-4t}}{6}\begin{pmatrix} 2 & 3 & 1 \\ -8 & -12 & -4 \\ 32 & 48 & 16 \end{pmatrix}$$

$$\int_0^t \mathrm{e}^{-\boldsymbol{A}t}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\mathrm{d}t = \frac{\mathrm{e}^t-1}{3}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} + \frac{\mathrm{e}^{2t}-1}{4}\begin{pmatrix} -1 \\ 2 \\ 4 \end{pmatrix} + \frac{\mathrm{e}^{4t}-1}{24}\begin{pmatrix} 1 \\ -4 \\ 16 \end{pmatrix}$$

因此所求常系数线性微分方程组的初值问题解是

$$y(t) = (1,0,0)\mathrm{e}^{\boldsymbol{A}t}\int_0^t \mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{f}(t)\mathrm{d}t = \frac{\mathrm{e}^t-1}{3} - \frac{\mathrm{e}^{2t}-1}{4} + \frac{\mathrm{e}^{4t}-1}{24}$$

定义 6.4.1 设 $\boldsymbol{A}$ 是 n 阶常数矩阵，如果对于任意的 t_0，初值问题

$$\begin{cases} \dfrac{\mathrm{d}x_1(t)}{\mathrm{d}t} = a_{11}x_1(t) + a_{12}x_2(t) + \cdots + a_{1n}x_n(t) \\ \dfrac{\mathrm{d}x_2(t)}{\mathrm{d}t} = a_{21}x_1(t) + a_{22}x_2(t) + \cdots + a_{2n}x_n(t) \\ \quad\vdots \\ \dfrac{\mathrm{d}x_n(t)}{\mathrm{d}t} = a_{n1}x_1(t) + a_{n2}x_2(t) + \cdots + a_{nn}x_n(t) \\ \boldsymbol{x}(t_0) = (c_1,c_2,\cdots,c_n)^{\mathrm{T}} \end{cases} \tag{6.4.10}$$

的解 $\boldsymbol{x}(t)$ 满足

$$\lim_{t\to+\infty}\boldsymbol{x}(t) = 0 \tag{6.4.11}$$

则称微分方程的解 $\boldsymbol{x}(t)$ 是渐近稳定的.

微分方程解的渐近稳定性是系统与控制理论的基本问题. 由于初值问题（6.4.10）的解

$$\boldsymbol{x}(t) = \mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}(t_0)$$

容易推出如下结论.

定理 6.4.1　对任意的 $\boldsymbol{x}(t_0)$，初值问题（6.4.10）的解 $\boldsymbol{x}(t)$ 渐近稳定的充分必要条件是 $\boldsymbol{A}$ 的特征值都有负实部.

定义 6.4.2　设 $\boldsymbol{A}$ 是 n 阶矩阵，如果 $\boldsymbol{A}$ 的特征值都有负实部，则称 $\boldsymbol{A}$ 是稳定矩阵.

作为矩阵函数的另一个应用，我们讨论线性系统的能控性与能观性. 我们只就连续型的线性定常系统进行讨论，进一步的讨论，读者可阅读有关文献.

设连续型的线性定常系统为

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{Ax}(t) + \boldsymbol{Bu}(t) & (6.4.12) \\ \boldsymbol{y}(t) = \boldsymbol{Cx}(t) + \boldsymbol{Du}(t) & (6.4.13) \end{cases}$$

其中 $\boldsymbol{A}$，$\boldsymbol{B}$，$\boldsymbol{C}$，$\boldsymbol{D}$ 均为常数矩阵．系统矩阵 $\boldsymbol{A}$ 是 $n\times n$ 型，输入矩阵 $\boldsymbol{B}$ 是 $n\times m$ 的，输出矩阵 $\boldsymbol{C}$ 是 $p\times n$ 的，又矩阵 $\boldsymbol{D}$ 是 $p\times m$ 的．状态向量 $\boldsymbol{x}(t)$ 是 n 维列向量，输入向量 $\boldsymbol{u}(t)$ 与输出向量 $\boldsymbol{y}(t)$ 分别是 m 维、p 维列向量．这个系统简称为系统（$\boldsymbol{A}$，$\boldsymbol{B}$，$\boldsymbol{C}$）.

定义 6.4.3　对于一个线性定常系统，若在某个有限时间区间 $[0,\ t_1]$ 内存在着输入 $\boldsymbol{u}(t)$，能使系统从初始状态 $\boldsymbol{x}(0)=\boldsymbol{x}_0$ 转移到 $\boldsymbol{x}(t_1)=0$，则称此状态 $\boldsymbol{x}_0$ 是能控的；若系统的所有状态部是能控的，则称此系统是完全能控的.

定理 6.4.2　系统（$\boldsymbol{A}$，$\boldsymbol{B}$，$\boldsymbol{C}$）是完全能控的充要条件是 n 阶对称矩阵

$$W_c(0,t_1) = \int_0^{t_1} \mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{BB}^{\mathrm{T}}\mathrm{e}^{-\boldsymbol{A}^{\mathrm{T}}\tau}\mathrm{d}\tau$$

非奇异.

证　充分性．设 $\boldsymbol{W}_c(0,\ t_1)$ 非奇异．对方程从 0 到 t_1 积分得

$$\boldsymbol{x}(t_1) = \mathrm{e}^{\boldsymbol{A}t_1}\boldsymbol{x}(0) + \int_0^{t_1} \mathrm{e}^{\boldsymbol{A}(t_1-\tau)}\boldsymbol{Bu}(\tau)\mathrm{d}\tau$$

令

$$\boldsymbol{u}(t) = -\boldsymbol{B}^{\mathrm{T}}\mathrm{e}^{-\boldsymbol{A}^{\mathrm{T}}t}\boldsymbol{W}_c^{-1}(0,t_1)\boldsymbol{x}(0)$$

则有

$$\begin{aligned} \boldsymbol{x}(t_1) &= \mathrm{e}^{\boldsymbol{A}t_1}\boldsymbol{x}(0) - \mathrm{e}^{\boldsymbol{A}t_1}\left(\int_0^{t_1} \mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{BB}^{\mathrm{T}}\mathrm{e}^{-\boldsymbol{A}^{\mathrm{T}}\tau}\mathrm{d}\tau\right)\boldsymbol{W}_c^{-1}(0,t_1)\boldsymbol{x}(0) \\ &= \mathrm{e}^{\boldsymbol{A}t_1}\boldsymbol{x}(0) - \mathrm{e}^{\boldsymbol{A}t_1}\boldsymbol{W}_c(0,t_1)\boldsymbol{W}_c^{-1}(0,t_1)\boldsymbol{x}(0) = 0 \end{aligned}$$

这说明在控制输入 $\boldsymbol{u}(t)$ 作用下，能使系统从 $\boldsymbol{x}(0)=\boldsymbol{x}_0$ 转移到 $\boldsymbol{x}(t_1)=0$.

必要性．用反证法．若系统是完全能控的，但 $\boldsymbol{W}_c(0,t_1)$ 是奇异的，则将引出矛盾.

因 $\boldsymbol{W}_c(0,t_1)$ 是奇异的，则必有非零向量 $\boldsymbol{\alpha}=(\alpha_1,\ \alpha_2,\ \cdots,\ \alpha_n)^{\mathrm{T}}$，使对任意时刻 $t=t_1\geqslant 0$ 下式成立

$$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{W}_c(0,t_1)\boldsymbol{\alpha} = 0$$

即

$$\int_0^{t_1} \boldsymbol{\alpha}^{\mathrm{T}}\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{BB}^{\mathrm{T}}\mathrm{e}^{-\boldsymbol{A}^{\mathrm{T}}\tau}\boldsymbol{\alpha}\,\mathrm{d}\tau = 0$$

故对任意时刻 t，有

$$\boldsymbol{\alpha}^{\mathrm{T}}\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{B} = 0,\quad t\geqslant 0$$

现系统是完全能控的，故必存在某个 $\boldsymbol{u}(t)$，使其作用于系统上，使得 $\boldsymbol{x}(t_1)=0$，将它代入

$$x(t_1)=e^{At_1}x(0)+\int_0^{t_1}e^{A(t_1-\tau)}Bu(\tau)d\tau$$

得

$$-x(0)=\int_0^{t_1}e^{-A\tau}Bu(\tau)d\tau$$

从而

$$-\alpha^Tx(0)=\alpha^T\int_0^{t_1}e^{-A\tau}Bu(\tau)d\tau=\int_0^{t_1}\alpha^Te^{-A\tau}Bu(\tau)d\tau=0$$

由于 $x(0)$ 是任意的，取 $x(0)=\alpha$，则有 $\alpha^T\alpha=0$. 这与 α 是非零向量相矛盾，故 $W_c(0,t_1)$ 非奇异.

用定理 6.4.2 来判别系统的能控性是不方便的，但在理论上有其重要性．下面的定理提供了个较简单的判别准则.

定理 6.4.3 系统（A，B，C）是完全能控的充要条件是 $n\times mn$ 矩阵

$$W_c=[B,AB,A^2B,\cdots,A^{n-1}B]$$

的秩为 n. W_c 称为能控性矩阵.

证 这里只给出充分性的证明. 如果系统不是完全能控的，则由定理 6.4.2 知，矩阵 $W_c(0,t_1)$ 是奇异的，由定理 6.4.2 的必要性证明中可知，必有非零向量 $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_n)^T$，使对任意时刻 t_1，有

$$\alpha^Te^{-At}B=0$$

上式两边逐次求微商，可得

$$\alpha^T(-A)^ke^{-At}B=0,\quad (k=1,2,\cdots,n-1)$$

令 $t=0$，得 $\alpha^TA^kB=0$，在式 $\alpha^Te^{-At}B=0$ 中取 $t=0$，得 $\alpha^TB=0$，故对于 $k=0$，1，2，…，$n-1$，都有 $\alpha^TA^kB=0$，即 $\alpha^T[B, AB, A^2B, \cdots, A^{n-1}B]=0$，这与矩阵 W_c 的秩为 n 相矛盾，从而证明了系统是完全能控的.

【例 6.4.5】 已知

$$A=\begin{pmatrix}1&3&2\\0&2&0\\0&1&3\end{pmatrix},B=\begin{pmatrix}2&1\\1&1\\-1&-1\end{pmatrix}$$

试判别系统（A，B，C）是否完全能控。

解 这是一个具有两个输入，三个状态变量的系统.

$$W_c=[B,AB,A^2B]=\begin{pmatrix}2&1&3&2&5&4\\1&1&2&2&4&4\\-1&-1&-2&-2&-4&-4\end{pmatrix}\sim\begin{pmatrix}2&1&3&2&5&4\\1&1&2&2&4&4\\0&0&0&0&0&0\end{pmatrix}$$

由于 $\mathrm{rank}(W_c)=2<3$，从而系统不是完全能控的.

下面我们来考虑能观测性问题.

定义 6.4.4 对于一个线性定常系统，若在有限时间区间 $[0, t_1]$ 内，能通过观测系统的输入 $y(t)$ 而唯一地确定任意的初始向量 $x(0)$，则称此系统是完全能观测的，或者说对每一个态 $x(0)$ 是能观测的.

定理 6.4.4 系统（A，B，C）是完全能观测的充要条件是 n 阶对称矩阵

$$M(0,t_1)=\int_0^{t_1}e^{A^T\tau}C^TCe^{A\tau}d\tau$$

为非奇异矩阵.

证

$$\begin{aligned} y(t) &= Cx(t) + Du(t) \\ &= Ce^{At}x(0) + C\int_0^{t_1} e^{A(t-\tau)}Bu(\tau)d\tau + Du(t) \end{aligned}$$

令

$$\eta(t) = y(t) - C\int_0^{t_1} e^{A(t-\tau)}Bu(\tau)d\tau - Du(t)$$

则有

$$Ce^{At_1}x(0) = \eta(t)$$

上式两边左乘 $e^{A^Tt}C^T$ 后，从 0 到 t_1 进行积分，可得

$$\int_0^{t_1} e^{A^Tt}C^TCe^{At}dtx(0) = \int_0^{t_1} e^{A^Tt}C^T\eta(t)dt$$

即

$$M(0,t_1)x(0) = \int_0^{t_1} e^{A^Tt}C^T\eta(t)dt$$

若 $M(0, t_1)$ 为非奇异矩阵，则

$$x(0) = M^{-1}(0,t_1)\int_0^{t_1} e^{A^Tt}C^T\eta(t)dt$$

即 $x(0)$ 能唯一确定. 由此可知，若 $M(0, t_1)$ 为非奇异，则系统是完全能观测的.

必要性. 若系统是完全能观测的. 现用反证法证明. 若 $M(0, t_1)$ 是奇异矩阵，必有非零向量 $\alpha=(\alpha_1, \alpha_2, \cdots, \alpha_n)^T$，使对任意时刻 $t=t_1\geqslant 0$ 下式成立.

$$\alpha^TM(0,t)\alpha = 0$$

即

$$\alpha^T\left(\int_0^{t_1} e^{A^T\tau}C^TCe^{A\tau}d\tau\right)\alpha = 0.$$

这表明对任意 $t\geqslant 0$ 时，有

$$Ce^{At}\alpha = 0$$

若取 $\alpha = x(0) \neq 0$，则对任意 $t\geqslant 0$，有

$$Ce^{At}x(0) = 0$$

但是

$$Ce^{At_1}x(0) = \eta(t)$$

这说明不能唯一确定. 而这与系统是完全能观测的假设相矛盾，定理证毕.

由于 $x(0)$ 是任意的，取 $x(0)=\alpha$，则有 $\alpha^T\alpha=0$. 这与 α 是非零向量相矛盾，故 $W_c(0,t_1)$ 非奇异.

定理 6.4.5 系统 (A, B, C) 是完全能观测的充要条件是 $pn\times n$ 矩阵

$$W_0 = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix}$$

的秩为 n，矩阵 W_0 叫做能观测性矩阵.

【例 6.4.6】　若某系统的状态方程与输出方程为

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \begin{pmatrix} -2 & 1 \\ -3 & -1 \end{pmatrix}\boldsymbol{x}(t) + \begin{pmatrix} 2 \\ 4 \end{pmatrix}\boldsymbol{u}(t)$$

$$\boldsymbol{y}(t) = \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix}\boldsymbol{x}(t)$$

试判定这系统的能观测性.

解　它的能观测性矩阵为

$$\boldsymbol{W}_0 = \begin{bmatrix} 2 & 1 \\ 1 & 5 \\ -5 & 0 \\ -13 & -1 \end{bmatrix}$$

显然它的秩等于 2，故系统是能观测的.

习　题　6

6.1　试讨论下列幂级数 $\sum_{n=1}^{\infty} \frac{1}{n^2}\begin{pmatrix} 1 & 4 \\ -1 & -3 \end{pmatrix}^n$ 的收敛性.

(1) $\sum_{n=1}^{\infty} \frac{n}{6^n}\begin{pmatrix} 1 & -8 \\ -2 & 1 \end{pmatrix}^n$；(2) $\sum_{n=1}^{\infty} \frac{1}{n^2}\begin{pmatrix} 1 & 4 \\ -1 & -3 \end{pmatrix}^n$；(3) $\sum_{n=1}^{\infty} \frac{1}{n^2}\begin{pmatrix} 1 & 7 \\ -1 & -3 \end{pmatrix}^n$.

6.2　求如下矩阵幂级数的和：

(1) $\sum_{n=1}^{\infty} \begin{pmatrix} 0.3 & -0.6 \\ 0.4 & 0.5 \end{pmatrix}^n$；(2) $\sum_{n=1}^{\infty} \frac{n}{10^n}\begin{pmatrix} -1 & 1 \\ -4 & 3 \end{pmatrix}^n$.

6.3　证明：若 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，$\rho(\boldsymbol{A})<1$，则 $\sum_{k=0}^{\infty} k\boldsymbol{A}^k = \boldsymbol{A}(\boldsymbol{E}-\boldsymbol{A})^{-2}$.

6.4　求下列三类矩阵的矩阵函数 $\cos\boldsymbol{A}$，$\sin\boldsymbol{A}$ 和 $\mathrm{e}^{\boldsymbol{A}}$.

(1) 当 $\boldsymbol{A}$ 为幂等矩阵 $\boldsymbol{A}^2=\boldsymbol{A}$ 时；

(2) 当 $\boldsymbol{A}$ 为对合矩阵 $\boldsymbol{A}^2=\boldsymbol{E}$ 时；

(3) 当 $\boldsymbol{A}$ 为幂零矩阵 $\boldsymbol{A}^2=0$ 时.

6.5　若 $\boldsymbol{A}$ 的特征值的实部全为负，证明：$\lim_{t\to+\infty} \mathrm{e}^{\boldsymbol{A}t}=0$.

6.6　设 4 阶矩阵 $\boldsymbol{A}$ 的特征值为 π，$-\pi$，0，0，求 $\sin\boldsymbol{A}$，$\cos\boldsymbol{A}$.

6.7　计算 $\mathrm{e}^{\boldsymbol{A}}$，$\sin\boldsymbol{A}$ 和 $\boldsymbol{A}^{100}-20\boldsymbol{A}^5+3\boldsymbol{A}-\boldsymbol{E}$，其中：

(1) $\begin{bmatrix} 4 & 5 & -2 \\ -2 & -2 & 1 \\ -1 & -1 & 1 \end{bmatrix}$；　(2) $\begin{bmatrix} 3 & 7 & -3 \\ -2 & -5 & 2 \\ -4 & -10 & 3 \end{bmatrix}$；

(3) $\begin{bmatrix} 1 & 1 & 2 \\ -5 & -5 & -10 \\ 4 & 4 & 8 \end{bmatrix}$；　(4) $\begin{bmatrix} 1 & -2 & -4 \\ -2 & 4 & -2 \\ -4 & -2 & 1 \end{bmatrix}$；

(5) $\begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$；　(6) $\begin{bmatrix} 2 & -1 & 2 \\ 5 & -3 & 3 \\ -1 & 0 & -2 \end{bmatrix}$.

6.8 设$A=\begin{pmatrix}9 & -6 & -2\\18 & -12 & -3\\18 & -9 & -6\end{pmatrix}$，求 e^A 的 Jordan 标准形.

6.9 设$A\in C^{n\times n}$，证明：

(1) $|e^A|=e^{\mathrm{tr}(A)}$；

(2) 对A的某一范数$\|\cdot\|$，若$\|E\|=1$，则$\|e^A\|\leqslant e^{\|A\|}$.

6.10 设 5 阶实矩阵A满足$A^3(A-2E)=0$，rank $(A)=3$，求$\|\sin A\|_2$，$\|\cos A\|_F$.

6.11 设$A=\begin{pmatrix}9 & -6 & -2\\18 & -12 & -3\\18 & -9 & -6\end{pmatrix}$，求微分方程$x'(t)=Ax(t)$的基础解系及满足初始条件$x(0)=(1,1,1)^T$的解.

6.12 设$A=\begin{pmatrix}1 & 0 & 0\\1 & 1 & 1\\1 & -1 & 3\end{pmatrix}$，$b(t)=\begin{pmatrix}1\\2\\t\end{pmatrix}$，求微分方程$x'(t)=Ax(t)+b(t)$满足初始条件$x(0)=(1,1,1)^T$的解.

第七章　矩 阵 的 分 解

将矩阵分解为具有某种特性的因子之积，从以我们所熟悉的 Gauss 消去法为依据而导出的 LU 分解，到 20 世纪六七十年代以 Givens 和 Householder 变换发展起来的 QR 分解，在矩阵理论的研究与应用中都具有十分重要的意义．这些特殊的分解式一方面反映了原矩阵的某些数值特征；另一方面，分解的方法与过程也为某些数值计算方法和理论分析提供了有效的工具．

§7.1　n 阶矩阵的 LU 分解

在线性代数中我们已经学过应用 Gauss 消去法求解 n 元线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$，

其中 $\boldsymbol{A}=(a_{ij})_{n\times n}$，$\boldsymbol{x}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$，$\boldsymbol{b}=(b_1, b_2, \cdots, b_n)^{\mathrm{T}}$．

Gauss 消去法的基本思路是将系数矩阵化为上三角形矩阵，或将增广矩阵化为上阶梯形矩阵，而后回代求解．

现在应用所谓选主元素法来实施 Gauss 消去法的消去过程，至于回代过程我们不做讨论．

设 $\boldsymbol{A}^{(0)}=\boldsymbol{A}$，记 $\boldsymbol{A}$ 的 k 阶顺序主子式为 $\Delta_k(k=1, 2, \cdots, n)$．

如果 $\Delta_1=a_{11}^{(0)}\neq 0$，令 $c_{i1}=\dfrac{a_{i1}^{(0)}}{a_{11}^{(0)}}(i=2, 3, \cdots, n)$，构造 Frobenius 矩阵

$$\boldsymbol{L}_1=\begin{pmatrix}1 & 0 & \cdots & 0\\ c_{21} & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ c_{n1} & 0 & \cdots & 1\end{pmatrix},\quad \boldsymbol{L}_1^{-1}=\begin{pmatrix}1 & 0 & \cdots & 0\\ -c_{21} & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ -c_{n1} & 0 & \cdots & 1\end{pmatrix}$$

则

$$\boldsymbol{L}_1^{-1}\boldsymbol{A}^{(0)}=\begin{pmatrix}a_{11}^{(0)} & a_{12}^{(0)} & \cdots & a_{1n}^{(0)}\\ 0 & a_{22}^{(1)} & \cdots & a_{2n}^{(1)}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & a_{n2}^{(1)} & \cdots & a_{nn}^{(1)}\end{pmatrix}=\boldsymbol{A}^{(1)} \tag{7.1.1}$$

因此，在 $\boldsymbol{A}^{(0)}$ 的第一列中除主元素 $a_{11}^{(0)}$ 外，其余元素均被化为零．式 (7.1.1) 即为 $\boldsymbol{A}^{(0)}=\boldsymbol{L}_1\boldsymbol{A}^{(1)}$，由于倍加变换不改变矩阵行列式的值，所以由 $\boldsymbol{A}^{(1)}$ 得到 $\boldsymbol{A}$ 的二阶顺序主子式为 $\Delta_2=a_{11}^{(0)}a_{22}^{(1)}$．

如果 $\Delta_2=a_{11}^{(0)}a_{22}^{(1)}\neq 0$，则必有 $a_{22}^{(1)}\neq 0$．令 $c_{i2}=\dfrac{a_{i2}^{(1)}}{a_{22}^{(1)}}(i=3, 4, \cdots, n)$，构造 Frobenius 矩阵

$$L_2=\begin{pmatrix}1&0&0&\cdots&0\\0&1&0&\cdots&0\\0&c_{32}&1&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&c_{n2}&0&\cdots&1\end{pmatrix},\quad L_2^{-1}=\begin{pmatrix}1&0&0&\cdots&0\\0&1&0&\cdots&0\\0&-c_{32}&1&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&-c_{n2}&0&\cdots&1\end{pmatrix}$$

则

$$L_2^{-1}A^{(1)}=\begin{pmatrix}a_{11}^{(0)}&a_{12}^{(0)}&a_{13}^{(0)}&\cdots&a_{14}^{(0)}\\0&a_{22}^{(1)}&a_{23}^{(1)}&\cdots&a_{2n}^{(1)}\\0&0&a_{33}^{(2)}&\cdots&a_{3n}^{(2)}\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&a_{n3}^{(2)}&\cdots&a_{nn}^{(2)}\end{pmatrix}=A^{(2)}\tag{7.1.2}$$

因此，在 $A^{(2)}$ 的前两列中，主元素以下的元素均已化为零. 式（7.1.2）即为 $A^{(1)}=L_2A^{(2)}$，由于倍加变换不改变矩阵行列式的值，所以由 $A^{(2)}$ 得到 A 的三阶顺序主子式为 $\Delta_3=a_{11}^{(0)}a_{22}^{(1)}a_{33}^{(2)}$.

依此类推，直到第 $r-1$ 步，得到

$$A^{(r-1)}=\begin{pmatrix}a_{11}^{(0)}&\cdots&a_{1,r-1}^{(0)}&a_{1r}^{(0)}&\cdots&a_{1n}^{(0)}\\\vdots&\ddots&\vdots&\vdots&\cdots&\vdots\\0&\cdots&a_{r-1,r-1}^{(r-2)}&a_{r-1,r}^{(r-2)}&\cdots&a_{r-1,n}^{(r-2)}\\0&\cdots&0&a_{rr}^{(r-1)}&\cdots&a_{rn}^{(r-1)}\\\vdots&&\vdots&\vdots&\ddots&\vdots\\0&\cdots&0&a_{nr}^{(r-1)}&\cdots&a_{nn}^{(r-1)}\end{pmatrix}$$

$$\Delta_r=a_{11}^{(0)}a_{22}^{(1)}\cdots a_{r-1,r-1}^{(r-2)}a_{rr}^{(r-1)}$$

如果 $\Delta_r\neq0$，则必有 $a_{rr}^{(r-1)}\neq0$. 令 $c_{ir}=\dfrac{a_{ir}^{(r-1)}}{a_{rr}^{(r-1)}}$ $(i=r+1,\ r+2,\ \cdots,\ n)$，构造 Frobenius 矩阵

$$L_r=\begin{pmatrix}1&&&&\\&\ddots&&&\\&&1&&\\&&c_{r+1,r}&1&\\&&\vdots&\vdots&\ddots\\&&c_{nr}&0&\cdots&1\end{pmatrix},\quad L_r^{-1}=\begin{pmatrix}1&&&&\\&\ddots&&&\\&&1&&\\&&-c_{r+1,r}&1&\\&&\vdots&\vdots&\ddots\\&&-c_{nr}&0&\cdots&1\end{pmatrix}$$

计算

$$L_r^{-1}A^{(r-1)}=\begin{pmatrix}a_{11}^{(0)}&\cdots&a_{1r}^{(0)}&a_{1,r+1}^{(0)}&\cdots&a_{1n}^{(0)}\\\vdots&\ddots&\vdots&\vdots&\ddots&\vdots\\0&\cdots&a_{rr}^{(r-1)}&a_{r,r+1}^{(r-1)}&\cdots&a_{rn}^{(r-1)}\\0&\cdots&0&a_{r+1,r+1}^{(r)}&\cdots&a_{r+1,n}^{(r)}\\\vdots&\ddots&\vdots&\vdots&\ddots&\vdots\\0&\cdots&0&a_{n,r+1}^{(r)}&\cdots&a_{nn}^{(r)}\end{pmatrix}=A^{(r)}\tag{7.1.3}$$

而在$A^{(r)}$的前r列中，主元素以下的元素均已化为零. 式（7.1.3）即为$A^{(r-1)}=L_rA^{(r)}$.

由于倍加变换不改变矩阵行列式的值，所以由$A^{(r)}$得到A的$r+1$阶顺序主子式为$\Delta_{r+1}=a_{11}^{(0)}a_{22}^{(1)}\cdots a_{rr}^{(r-1)}a_{r+1,r+1}^{(r)}$.

依此类推，在第$n-1$步之后得到

$$A^{(n-1)}=\begin{pmatrix} a_{11}^{(0)} & a_{12}^{(0)} & a_{13}^{(0)} & \cdots & a_{14}^{(0)} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{2n}^{(1)} \\ 0 & 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & a_{n-1,n-1}^{(n-2)} & a_{n-1,n}^{(n-2)} \\ 0 & 0 & 0 & 0 & a_{nn}^{(n-1)} \end{pmatrix}$$

以上即为 Gauss 消元的过程，Gauss 消元过程能够进行到底的充分且必要条件是$a_{11}^{(0)}$，$a_{22}^{(1)}$，…，$a_{n-1,n-1}^{(n-2)}$均不等于零，即

$$\Delta_r = a_{11}^{(0)}a_{22}^{(1)}\cdots a_{r-1,r-1}^{(r-2)}a_{rr}^{(r-1)} \neq 0(r=1,2,\cdots,n-1) \tag{7.1.4}$$

由于 Gauss 顺序消元过程没有用到行、列的交换，所以条件（7.1.4）是合理的.

在条件（7.1.4）下，有$A=A^{(0)}=L_1A^{(1)}=L_1L_2A^{(2)}=\cdots=L_1L_2\cdots L_{n-1}A^{(n-1)}$

所以

$$L=L_1L_2\cdots L_{n-1}=\begin{pmatrix} 1 & & & & \\ c_{21} & 1 & & & \\ \vdots & \vdots & \ddots & & \\ c_{n-1,1} & c_{n-1,2} & \cdots & 1 & \\ c_{n1} & c_{n2} & \cdots & c_{n,n-1} & 1 \end{pmatrix}$$

是一个主对角线上的元素都是1的下三角矩阵，我们称之为单位下三角矩阵，而$A^{(n-1)}$是一个上三角矩阵. 令$U=A^{(n-1)}$，则$A=LU$.

定义 7.1.1 如果n阶矩阵A能够分解为一个下三角矩阵L和一个上三角矩阵U的乘积，则称其为三角分解或LU分解. 如果n阶矩阵A能够分解为$A=LDU$，其中L为单位下三角矩阵，D为对角矩阵，U为单位上三角矩阵，则称之为LDU分解.

在一个n阶矩阵A的LU分解中，若取一个主对角线上的元素都不为零的对角矩阵D，则$A=LU$就可以记作$A=LDD^{-1}U=\hat{L}\hat{U}$，其中$LD=\hat{L}$，$D^{-1}U=\hat{U}$仍分别为下三角矩阵与上三角矩阵，因此，$n$阶矩阵$A$的$LU$分解不唯一.

定理 7.1.1 矩阵$A=(a_{ij})_{n\times n}$的LDU分解式唯一的充分且必要条件为A的顺序主子式$\Delta_k\neq 0(k=1, 2, \cdots, n-1)$. 其中$L$是单位下三角矩阵，$U$是单位上三角矩阵，

$D=\mathrm{diag}(d_1, d_2, \cdots, d_n)$是对角矩阵，并且$d_k=\dfrac{\Delta_k}{\Delta_{k-1}}(\Delta_0=1, k=1, 2, \cdots, n)$.

推论 7.1.1 n阶非奇异矩阵A有LU分解的充分且必要条件为A的顺序主子式$\Delta_k\neq 0(k=1, 2, \cdots, n-1)$.

【例 7.1.1】 求矩阵$A=\begin{pmatrix} 2 & -1 & 3 \\ 1 & 2 & 1 \\ 2 & 4 & 3 \end{pmatrix}$的$LDU$分解.

解 因为$\Delta_1=2$，$\Delta_2=5$，所以A有唯一的LDU分解，构造 Frobenius 矩阵

$$L_1=\begin{pmatrix}1&0&0\\1/2&1&0\\1&0&1\end{pmatrix},L_1^{-1}=\begin{pmatrix}1&0&0\\-1/2&1&0\\-1&0&1\end{pmatrix}$$

则

$$L_1^{-1}A^{(0)}=\begin{pmatrix}2&-1&3\\0&5/2&-1/2\\0&5&-1\end{pmatrix}=A^{(1)}.$$

对 $A^{(1)}$ 构造 Frobenius 矩阵

$$L_2=\begin{pmatrix}1&0&0\\0&1&0\\0&2&1\end{pmatrix},L_1^{-1}=\begin{pmatrix}1&0&0\\0&1&0\\0&-2&1\end{pmatrix}$$

则

$$L_2^{-1}A^{(1)}=\begin{pmatrix}2&-1&3\\0&5/2&-1/2\\0&0&1\end{pmatrix}==\begin{pmatrix}2&0&0\\0&5/2&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&-1/2&3/2\\0&1&-1/5\\0&0&1\end{pmatrix}=A^{(2)}$$

因此，

$$L=L_1L_2=\begin{pmatrix}1&0&0\\1/2&1&0\\1&2&1\end{pmatrix}$$

所以

$$A=L_1L_2A^{(2)}=\begin{pmatrix}1&0&0\\1/2&1&0\\1&2&1\end{pmatrix}\begin{pmatrix}2&0&0\\0&5/2&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&-1/2&3/2\\0&1&-1/5\\0&0&1\end{pmatrix}=LDU.$$

在计算机编程中为节省内存，同时为直接求 L，可按如下计算方法.

$$A=\begin{pmatrix}2&-1&3\\1&2&1\\2&4&2\end{pmatrix}\sim\begin{pmatrix}2&-1&3\\\frac{1}{2}&\frac{5}{2}&-\frac{1}{2}\\1&5&-1\end{pmatrix}\sim\begin{pmatrix}2&-1&3\\\frac{1}{2}&\frac{5}{2}&-\frac{1}{2}\\1&2&1\end{pmatrix}$$

因此

$$A=\begin{pmatrix}1&0&0\\\frac{1}{2}&1&0\\1&2&1\end{pmatrix}\begin{pmatrix}2&-1&3\\0&\frac{5}{2}&-\frac{1}{2}\\0&0&1\end{pmatrix}$$

$$=\begin{pmatrix}1&0&0\\\frac{1}{2}&1&0\\1&2&1\end{pmatrix}\begin{pmatrix}2&0&0\\0&\frac{5}{2}&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&-\frac{1}{2}&\frac{3}{2}\\0&1&-\frac{1}{5}\\0&0&1\end{pmatrix}$$

由于矩阵 A 的 LU 分解与 LDU 分解都要求 A 的前 $n-1$ 阶顺序主子式不为零. 如果这个条件不满足，则可以对矩阵 A 右（左）乘以置换矩阵 P，将矩阵 A 的列（或行）重新排列顺

序，使排列后的矩阵的前 $n-1$ 阶顺序主子式不为零．因此有以下的定理．

定理 7.1.2 设 $\boldsymbol{A}$ 是 n 阶非奇异矩阵，则存在置换矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{PA}$ 的 n 阶顺序主子式不为零．

推论 7.1.2 设 $\boldsymbol{A}$ 是 n 阶非奇异矩阵，则存在置换矩阵 $\boldsymbol{P}$，使

$$\boldsymbol{PA} = \boldsymbol{L}\hat{\boldsymbol{U}} = \boldsymbol{LDU} \tag{7.1.5}$$

其中 $\boldsymbol{L}$ 是单位下三角矩阵，$\hat{\boldsymbol{U}}$ 是上三角矩阵，$\boldsymbol{U}$ 是单位上三角矩阵，$\boldsymbol{D}$ 是对角矩阵．

对于线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 而言，如果其系数矩阵 $\boldsymbol{A}$ 为非奇异矩阵，并且 $\Delta_k \neq 0$ $(k=1, 2, \cdots, n-1)$，则存在三角分解 $\boldsymbol{A}=\boldsymbol{LU}$．

这时，$\boldsymbol{Ax}=\boldsymbol{b}$ 就与方程组 $\begin{cases}\boldsymbol{Ly}=\boldsymbol{b}\\ \boldsymbol{Ux}=\boldsymbol{y}\end{cases}$ 等价，而方程组 $\begin{cases}\boldsymbol{Ly}=\boldsymbol{b}\\ \boldsymbol{Ux}=\boldsymbol{y}\end{cases}$ 中的两个子方程组很容易求解，这就是解线性方程组的三角分解法．

将矩阵 $\boldsymbol{A}$ 分解为一个单位下三角矩阵与一个上三角矩阵的乘积的分解法叫做 Doolittle 分解．如同 Gauss 顺序消去法不能消元到底一样，矩阵 $\boldsymbol{A}$ 也不一定存在 Doolittle 分解．

定理 7.1.3 n 阶矩阵 $\boldsymbol{A}$ $(n\geqslant 2)$ 有唯一的 Doolittle 分解的充分且必要条件是 $\boldsymbol{A}_1$，$\boldsymbol{A}_2$，$\cdots$，$\boldsymbol{A}_{n-1}$ 为非奇异矩阵，其中 $\boldsymbol{A}_k$ $(k=1, 2, \cdots, n-1)$ 是 $\boldsymbol{A}$ 的 k 阶顺序主子矩阵．

证 （充分性）应用数学归纳法．设 $\boldsymbol{A}_1$，$\boldsymbol{A}_2$，$\cdots$，$\boldsymbol{A}_{n-1}$ 为非奇异矩阵，显然有 $\boldsymbol{A}_1 = a_{11} \neq 0$．设 $\boldsymbol{A}_k$ $(k\leqslant n-1)$ 可以分解为 $\boldsymbol{A}_k=\boldsymbol{L}_k\boldsymbol{U}_k$，其中 $\boldsymbol{L}_k$ 是单位下三角矩阵，$\boldsymbol{U}_k$ 是非奇异上三角矩阵．$\boldsymbol{A}$ 的 $k+1$ 阶顺序主子矩阵 $\boldsymbol{A}_{k+1}=\begin{pmatrix}\boldsymbol{A}_k & \boldsymbol{y}\\ \boldsymbol{x}^{\mathrm{T}} & \boldsymbol{a}\end{pmatrix}$，其中向量 $\boldsymbol{x}^{\mathrm{T}}$，$\boldsymbol{y}$ 与元素 $\boldsymbol{a}$ 已知．

令

$$\begin{pmatrix}\boldsymbol{A}_k & \boldsymbol{y}\\ \boldsymbol{x}^{\mathrm{T}} & \boldsymbol{a}\end{pmatrix} = \begin{pmatrix}\boldsymbol{L}_k & 0\\ \boldsymbol{p}^{\mathrm{T}} & 1\end{pmatrix}\begin{pmatrix}\boldsymbol{U}_k & \boldsymbol{q}\\ 0 & \boldsymbol{u}\end{pmatrix}$$

则有

$$\boldsymbol{L}_k\boldsymbol{q} = \boldsymbol{y}, \boldsymbol{p}^{\mathrm{T}}\boldsymbol{U}_k = \boldsymbol{x}^{\mathrm{T}}, \boldsymbol{p}^{\mathrm{T}}\boldsymbol{q} + \boldsymbol{u} = \boldsymbol{a}$$

即

$$\boldsymbol{q} = \boldsymbol{L}_k^{-1}\boldsymbol{y}, \boldsymbol{p}^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{U}_k^{-1}, \boldsymbol{u} = \boldsymbol{a} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}_k^{-1}\boldsymbol{y}$$

记

$$\boldsymbol{L}_{k+1} = \begin{pmatrix}\boldsymbol{L}_k & 0\\ \boldsymbol{p}^{\mathrm{T}} & 1\end{pmatrix}, \boldsymbol{U}_{k+1} = \begin{pmatrix}\boldsymbol{U}_k & \boldsymbol{q}\\ 0 & \boldsymbol{u}\end{pmatrix}$$

则有 $\boldsymbol{A}_{k+1}=\boldsymbol{L}_{k+1}\boldsymbol{U}_{k+1}$，并且 $\boldsymbol{L}_{k+1}$ 是单位下三角矩阵，$\boldsymbol{U}_{k+1}$ 是上三角矩阵，只要 $k<n-1$，由于 $\boldsymbol{A}_{k+1}$ 是非奇异矩阵，所以 $\boldsymbol{U}_{k+1}$ 也是非奇异矩阵．

由归纳法，当 $k=n-1$ 时，有 $\boldsymbol{A}=\boldsymbol{A}_n=\boldsymbol{LU}$，其中 $\boldsymbol{L}$ 是单位下三角矩阵，$\boldsymbol{U}$ 是上三角矩阵．如果 $\boldsymbol{A}$ 是非奇异矩阵，则 $\boldsymbol{U}$ 也是非奇异矩阵；如果 $\boldsymbol{A}$ 是奇异矩阵，则 $\boldsymbol{U}$ 的主对角线上的元素 $u_{ii}\neq 0$ $(i=1, 2, \cdots, n-1)$，$u_{nn}=0$．

再设 $\boldsymbol{A}$ 有两种 Doolittle 分解 $\boldsymbol{A}=\boldsymbol{LU}=\boldsymbol{L}^*\boldsymbol{U}^*$．如果 $\boldsymbol{A}$ 是非奇异矩阵，由此式可得 $\boldsymbol{L}^{*-1}\boldsymbol{L}=\boldsymbol{U}^*\boldsymbol{U}^{-1}$，由于 $\boldsymbol{L}^{*-1}\boldsymbol{L}$ 是单位下三角矩阵，$\boldsymbol{U}^*\boldsymbol{U}^{-1}$ 是上三角矩阵，所以它们只能是单位矩阵，即 $\boldsymbol{L}^{*-1}\boldsymbol{L}=\boldsymbol{U}^*\boldsymbol{U}^{-1}=\boldsymbol{E}$，所以 $\boldsymbol{L}=\boldsymbol{L}^*$，$\boldsymbol{U}=\boldsymbol{U}^*$．如果 $\boldsymbol{A}$ 是奇异矩阵，则将式 $\boldsymbol{A}=\boldsymbol{LU}=\boldsymbol{L}^*\boldsymbol{U}^*$ 写成分块形式

$$\begin{pmatrix} L_{n-1} & 0 \\ r^{\mathrm{T}} & 1 \end{pmatrix}\begin{pmatrix} U_{n-1} & s \\ 0 & 0 \end{pmatrix}=\begin{pmatrix} L_{n-1}^{*} & 0 \\ r^{*\mathrm{T}} & 1 \end{pmatrix}\begin{pmatrix} U_{n-1}^{*} & s^{*} \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} L_{n-1}U_{n-1} & L_{n-1}s \\ r^{\mathrm{T}}U_{n-1} & r^{\mathrm{T}}s \end{pmatrix}=\begin{pmatrix} L_{n-1}^{*}U_{n-1}^{*} & L_{n-1}^{*}s^{*} \\ r^{*\mathrm{T}}U_{n-1}^{*} & r^{*\mathrm{T}}s^{*} \end{pmatrix}$$

其中L_{n-1}，U_{n-1}，L_{n-1}^{*}，U_{n-1}^{*}都是非奇异矩阵，类似的有

$$L_{n-1}=L_{n-1}^{*},U_{n-1}=U_{n-1}^{*},r=r^{*},s=s^{*}$$

所以，$L=L^{*}$，$U=U^{*}$.

（必要性）设A有唯一的Doolittle分解$A=LU$. 当A是非奇异矩阵时，L与U都是非奇异矩阵；如果A是奇异矩阵，则必有$u_{nn}=0$，$u_{ii}\neq 0(i=1, 2, \cdots, n-1)$，否则，要么$A$的Doolittle分解不存在，要么这种分解虽然存在但不唯一. 因此，不论A是否是奇异矩阵，L与U的k阶顺序主子矩阵L_k与$U_k(k=1, 2, \cdots, n-1)$都是非奇异矩阵. 而$A_k=L_kU_k$ $(k=1, 2, \cdots, n-1)$，所以A_1，A_2，$\cdots$，A_{n-1}为非奇异矩阵.

推论 7.1.3 顺序Gauss消去法的主元素$a_{kk}^{(k)}\neq 0(k=1, 2, \cdots, n-1)$的充分且必要条件是方程组的系数矩阵$A$的$k(k=1, 2, \cdots, n-1)$阶顺序主子矩阵是非奇异矩阵.

以下来看Doolittle分解如何进行. 设A的顺序主子矩阵A_1，A_2，$\cdots$，A_{n-1}为非奇异矩阵，且$L=(l_{ij})$，$U=(u_{ij})$. 因为$A=LU$，即

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}=\begin{pmatrix} 1 & & & \\ l_{21} & 1 & & \\ \vdots & \ddots & \ddots & \\ l_{n1} & \cdots & l_{n,n-1} & 1 \end{pmatrix}\begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & u_{nn} \end{pmatrix}$$

所以，$a_{1j}=u_{1j}$，$j=1, 2, \cdots, n$

$$a_{ij}=\begin{cases} \sum_{t=1}^{j} l_{it}u_{tj} & j<i \\ \sum_{t=1}^{i-1} l_{it}u_{tj}+u_{ij} & j\geqslant i \end{cases}$$

由此得到计算l_{ij}与u_{ij}的递推公式：$u_{1j}=a_{1j}$，$j=1, 2, \cdots, n$. 对于$i=2, 3, \cdots, n$，计算

$$\begin{cases} l_{ij}=(a_{ij}-\sum_{t=1}^{j-1} l_{it}u_{tj})/u_{jj} & j=1,2,\cdots,i-1 \\ u_{ij}=a_{ij}-\sum_{t=1}^{i-1} l_{it}u_{tj} & j=i,i+1,\cdots,n \end{cases}$$

当A为实对称正定矩阵时，$\Delta_k>0(k=1, 2, \cdots, n)$. 于是$A$有唯一的$LDU$分解，即$A=LDU$，其中$D=\mathrm{diag}(d_1, d_2, \cdots, d_n)$，且$d_i>0(i=1, 2, \cdots, n)$.

令$\tilde{D}=\mathrm{diag}(\sqrt{d_1}, \sqrt{d_2}, \cdots, \sqrt{d_n})$，则有$A=L\tilde{D}^2U$，由$A^{\mathrm{T}}=A$，因此$L\tilde{D}^2U=U^{\mathrm{T}}\tilde{D}^2L^{\mathrm{T}}$，再由分解的唯一性有$L=U^{\mathrm{T}}$，$U=L^{\mathrm{T}}$，所以

$$A=L\tilde{D}^2L^{\mathrm{T}}=LDL^{\mathrm{T}}$$

或者

$$A=L\tilde{D}^2L^{\mathrm{T}}=(L\tilde{D})(L\tilde{D})^{\mathrm{T}}=GG^{\mathrm{T}} \tag{7.1.6}$$

其中$G=L\tilde{D}$是下三角矩阵.

定义 7.1.2 实对称正定矩阵 $\boldsymbol{A}$ 的分解式（7.1.6）叫做实对称正定矩阵 $\boldsymbol{A}$ 的 Cholesky 分解（也叫做平方根分解，或对称三角分解）.

令 $\boldsymbol{G}=(g_{ij})$，由式（7.1.6）有

$$a_{ij}=g_{i1}g_{j1}+g_{i2}g_{j2}+\cdots+g_{ij}g_{jj},i>j$$
$$a_{ii}=g_{i1}^2+g_{i2}^2+\cdots+g_{ii}^2$$

由此得到计算 g_{ij} 的公式

$$g_{ij}=\begin{cases}\left(a_{ii}-\sum\limits_{k=1}^{i-1}g_{ik}^2\right)^{1/2}, & i=j\\ \dfrac{1}{g_{ii}}\left(a_{ij}-\sum\limits_{k=1}^{j-1}g_{ik}g_{jk}\right), & i>j\\ 0, & i<j\end{cases} \tag{7.1.7}$$

因为 $\boldsymbol{A}$ 的对角元素 $a_{ii}=\sum\limits_{j=1}^{i}g_{ij}^2$ ，所以 $|g_{ij}|\leqslant\sqrt{a_{ii}}$，$(j\leqslant i)$ 由此式可知 Cholesky 分解中的中间量 g_{ij} 得以控制.

将矩阵 $\boldsymbol{A}$ 分解为一个下三角矩阵和一个单位上三角矩阵之积的分解，叫做矩阵 $\boldsymbol{A}$ 的 Crout 分解.

定理 7.1.4 $n\times n$ 阶矩阵 $\boldsymbol{A}\,(n\geqslant 2)$ 有唯一的 Crout 分解的充分且必要条件是 $\boldsymbol{A}$ 的顺序主子式 $\boldsymbol{A}_1$，$\boldsymbol{A}_2$，…，$\boldsymbol{A}_{n-1}$ 为非奇异矩阵. 仍用 $\boldsymbol{A}=\boldsymbol{LU}$ 表示 $\boldsymbol{A}$ 的 Crout 分解，这里 $\boldsymbol{L}$ 是下三角矩阵，$\boldsymbol{U}$ 是单位上三角矩阵，$\boldsymbol{L}=(l_{ij})$，$\boldsymbol{U}=(u_{ij})$.

与定理 7.1.3 类似可得到计算 l_{ij} 与 u_{ij} 的递推公式

$$l_{i1}=a_{i1},i=1,2,\cdots,n \tag{7.1.8}$$

对于 $j=2$，3，…，n，计算

$$\begin{cases}u_{ij}=\left(a_{ij}-\sum\limits_{t=1}^{i-1}l_{it}u_{tj}\right)/l_{ii} & i=1,2,\cdots,j-1\\ l_{ij}=a_{ij}-\sum\limits_{t=1}^{j-1}l_{it}u_{tj} & i=j,j+1,\cdots,n\end{cases} \tag{7.1.9}$$

【例 7.1.2】 求矩阵 $\boldsymbol{A}=\begin{pmatrix}2 & -1 & 3\\ 1 & 2 & 1\\ 2 & 4 & 3\end{pmatrix}$ 的 Doolittle 分解与 Crout 分解.

解 由 Doolittle 分解的递推公式有

$$u_{11}=a_{11}=2,u_{12}=a_{12}=-1,u_{13}=a_{13}=3$$
$$l_{21}=\frac{a_{21}}{u_{11}}=\frac{1}{2},l_{31}=\frac{a_{31}}{u_{11}}=1$$
$$u_{22}=a_{22}-l_{21}u_{12}=\frac{5}{2},u_{23}=a_{23}-l_{21}u_{13}=-\frac{1}{2}$$
$$l_{32}=\frac{1}{u_{22}}(a_{32}-l_{31}u_{12})=2$$
$$u_{33}=a_{33}-l_{31}u_{13}-l_{32}u_{23}=1$$

所以 Doolittle 分解为

$$A=\begin{pmatrix}1&0&0\\1/2&1&0\\1&2&1\end{pmatrix}\begin{pmatrix}2&-1&3\\0&5/2&-1/2\\0&0&1\end{pmatrix}$$

再由 Crout 分解的递推公式有

$$l_{11}=a_{11}=2,l_{21}=a_{21}=1,l_{31}=a_{31}=2,u_{12}=\frac{a_{12}}{l_{11}}=-\frac{1}{2}$$

$$u_{13}=\frac{a_{13}}{l_{11}}=\frac{3}{2},l_{22}=a_{22}-l_{21}u_{12}=\frac{5}{2},l_{32}=a_{32}-l_{31}u_{12}=5$$

$$u_{23}=\frac{1}{l_{22}}(a_{23}-l_{21}u_{13})=-\frac{1}{5},l_{33}=a_{33}-l_{31}u_{13}-l_{32}u_{23}=1$$

所以 Crout 分解为

$$A=\begin{pmatrix}2&0&0\\1&5/2&0\\2&5&1\end{pmatrix}\begin{pmatrix}1&-1/2&3/2\\0&1&-1/5\\0&0&1\end{pmatrix}$$

【例 7.1.3】 求矩阵 $A=\begin{pmatrix}5&-2&0\\-2&3&-1\\0&-1&1\end{pmatrix}$ 的 Cholesky 分解.

解 易于验证矩阵 A 是实对称正定矩阵，由 Cholesky 分解的递推公式有

$$g_{11}=\sqrt{a_{11}}=\sqrt{5},g_{21}=\frac{a_{21}}{g_{11}}=-\frac{2}{\sqrt{5}},g_{31}=\frac{a_{31}}{g_{11}}=0$$

$$g_{22}=\sqrt{a_{22}-|g_{21}|^2}=\sqrt{\frac{11}{5}},g_{32}=\frac{1}{g_{22}}(a_{32}-g_{31}g_{21})=-\sqrt{\frac{5}{11}}$$

$$g_{33}=\sqrt{a_{33}-|g_{31}|^2-|g_{32}|^2}=\sqrt{\frac{6}{11}}$$

所以 Cholesky 分解为

$$A=\begin{pmatrix}\sqrt{5}&0&0\\-2/\sqrt{5}&\sqrt{11/5}&0\\0&-\sqrt{5/11}&\sqrt{6/11}\end{pmatrix}\begin{pmatrix}\sqrt{5}&-2/\sqrt{5}&0\\0&\sqrt{11/5}&-\sqrt{5/11}\\0&0&\sqrt{6/11}\end{pmatrix}$$

要记上述公式不易，在计算时可用如下方法：

(1) 对 A 进行 Dollitle 分解 $A=LU$；

(2) 选取 U 的对角线元素 d_1，d_2，…，d_n；

(3) 令 $G=L\cdot\mathrm{diag}\,(\sqrt{d_1}, \sqrt{d_2}, \cdots, \sqrt{d_n})$；

(4) 写出 Cholesky 分解 $A=GG^{\mathrm{T}}$.

【例 7.1.4】 求矩阵 $A=\begin{pmatrix}5&-2&0\\-2&3&-1\\0&-1&1\end{pmatrix}$ 的 Cholesky 分解.

解 $$A=\begin{pmatrix}5&-2&0\\-2&3&-1\\0&-1&1\end{pmatrix}\sim\begin{pmatrix}5&-2&0\\-\frac{2}{5}&\frac{11}{5}&-1\\0&-1&1\end{pmatrix}\sim\begin{pmatrix}5&-2&0\\-\frac{2}{5}&\frac{11}{5}&-1\\0&-\frac{5}{11}&\frac{6}{11}\end{pmatrix}$$

A 进行 Dollitle 分解 $A=LU$ 为

$$LU=\begin{pmatrix}1&0&0\\-\frac{2}{5}&1&0\\0&-\frac{5}{11}&1\end{pmatrix}\begin{pmatrix}5&-2&0\\0&\frac{11}{5}&-1\\0&0&\frac{6}{11}\end{pmatrix}$$

$$G=\begin{pmatrix}1&0&0\\-\frac{2}{5}&1&0\\0&-\frac{5}{11}&1\end{pmatrix}\begin{pmatrix}\sqrt{5}&&\\&\sqrt{\frac{11}{5}}&\\&&\sqrt{\frac{6}{11}}\end{pmatrix}=\begin{pmatrix}\sqrt{5}&0&0\\-\frac{2\sqrt{5}}{5}&\sqrt{\frac{11}{5}}&0\\0&-\sqrt{\frac{5}{11}}&\sqrt{\frac{6}{11}}\end{pmatrix}$$

所以 Cholesky 分解为

$$A=\begin{pmatrix}\sqrt{5}&0&0\\-2/\sqrt{5}&\sqrt{11/5}&0\\0&-\sqrt{5/11}&\sqrt{6/11}\end{pmatrix}\begin{pmatrix}\sqrt{5}&-2/\sqrt{5}&0\\0&\sqrt{11/5}&-\sqrt{5/11}\\0&0&\sqrt{6/11}\end{pmatrix}$$

LU 分解 MATLAB 计算以及在解方程中的应用如下.

LU 分解. 矩阵的 LU 分解就是将一个矩阵表示为一个交换下三角矩阵和一个上三角矩阵的乘积形式.

MATLAB 提供的 lu 函数用于对矩阵进行 LU 分解，其调用格式为

$$[L,U]=\text{lu}(X).$$

产生一个上三角阵 U 和一个变换形式的下三角阵 L（行交换），使之满足 $X=LU$. 注意，这里的矩阵 X 必须是方阵.

$$[L,U,P]=\text{lu}(X)$$

产生一个上三角阵 U 和一个下三角阵 L 以及一个置换矩阵 P，使之满足 $PX=LU$. 当然矩阵 X 同样必须是方阵.

实现 LU 分解后，线性方程组 $Ax=b$ 的解 $x=U\backslash(L\backslash b)$ 或 $x=U\backslash(L\backslash Pb)$，这样可以大大提高运算速度.

【例 7.1.5】 解如下线性方程组

$$\begin{cases}2x_1+x_2-5x_3+x_4=13\\x_1-5x_2+7x_4=-9\\2x_2+x_3-x_4=6\\x_1+6x_2-x_3-4x_4=0\end{cases}$$

解 用直接解法求解下列线性方程组. 命令如下.

```
A=[2,1,-5,1;1,-5,0,7;0,2,1,-1;1,6,-1,-4];
b=[13,-9,6,0]';
x=A\b
```

运行结果为：

```
x=
-66.5556
  25.6667
```

```
-18.7778
26.5556
```

用 LU 分解求解例 7.1.5 中的线性方程组. 命令如下.

```
A=[2,1,-5,1;1,-5,0,7;0,2,1,-1;1,6,-1,-4];
b=[13,-9,6,0]';
[L,U]=lu(A);
x=U\(L\b)
```

或采用 LU 分解的第 2 种格式，命令如下.

```
[L,U,P]=lu(A);
x=U\(L\P*b)
```

运行结果为：

```
x=
-66.5556
  25.6667
-18.7778
  26.5556
```

Cholesky 分解.

如果矩阵 X 是对称正定的，则 Cholesky 分解将矩阵 X 分解成一个下三角矩阵和上三角矩阵的乘积. 设上三角矩阵为 R，则下三角矩阵为其转置，即 $X=R^TR$. MATLAB 函数 chol（X）用于对矩阵 X 进行 Cholesky 分解，其调用格式为

$$R = \mathrm{chol}(X)$$

产生一个上三角阵 R，使 $R^TR=X$. 若 X 为非对称正定，则输出一个出错信息.

$$[R,\mathrm{p}] = \mathrm{chol}(X)$$

这个命令格式将不输出出错信息. 当 X 为对称正定的，则 p=0，R 与上述格式得到的结果相同；否则 p 为一个正整数. 如果 X 为满秩矩阵，则 R 为一个阶数为 q=p-1 的上三角阵，且满足 $R^TR=X$（1：q，1：q）.

实现 Cholesky 分解后，线性方程组 $Ax=b$ 变成 $R^TRx=b$，所以 $x=R\backslash(R^T\backslash b)$.

Doolittle 分解在 MATLAB 中没有直接的命令，我们可以建立如下 m 函数.

```
function X=myDoolittle(A)
[n,m]=size(A);
if n~=m-1
    'error'
    return
end
for i=2:n
    A(i,1)=A(i,1)/A(1,1);
end
for k=2:n-1
    for j=k:n
      A(k,j)=A(k,j)-sum(A(k,1:k-1)*A(1:k-1,j));
```

```
        end
        for i=k+1:n
        A(i,k)=(A(i,k)-sum(A(i,1:k-1)*A(1:k-1,k)))/A(k,k);
        end
    end
    A(n,n)=A(n,n)-sum(A(n,1:n-1)*A(1:n-1,n));
        if A(n,n)==0
          'error'
          return
        end
    for k=2:n
        A(k,n+1)=A(k,n+1)-sum(A(k,1:k-1)*A(1:k-1,n+1));
    end
    X(n)=A(n,n+1)/A(n,n);
    for k=n-1:-1:1
        X(k)=(A(k,n+1)-sum(A(k,k+1:n).*X(k+1:n)))/A(k,k);
    end
```

§7.2 矩阵的 QR 分解

本节首先介绍 Givens 矩阵与 Householder 矩阵，并讨论可逆矩阵 **A** 的 **QR** 分解.

1. Givens 矩阵与 Givens 变换

在二维平面中，笛卡尔直角坐标系的旋转变换为 $\boldsymbol{y}=\begin{pmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{pmatrix}\boldsymbol{x}=\boldsymbol{T}\boldsymbol{x}$，这是一个正交变换，所以 $\boldsymbol{T}$ 是正交矩阵，而 $\det\boldsymbol{T}=1$.

一般的，对于 n 维欧氏空间而言，取定一个标准正交基 $\boldsymbol{e}_1$，$\boldsymbol{e}_2$，…，$\boldsymbol{e}_n$，则在平面 $[\boldsymbol{e}_i, \boldsymbol{e}_j]$ 中，旋转变换定义如下.

定义 7.2.1 设实数 c 与 s 满足 $c^2+s^2=1$，则称矩阵

$$\boldsymbol{T}_{ij}=\begin{pmatrix}
1 & & & & & & & & & & \\
& \ddots & & & & & & & & & \\
& & 1 & & & & & & & & \\
& & & c & & & & s & & & \\
& & & & 1 & & & & & & \\
& & & & & \ddots & & & & & \\
& & & & & & 1 & & & & \\
& & & -s & & & & c & & & \\
& & & & & & & & 1 & & \\
& & & & & & & & & \ddots & \\
& & & & & & & & & & 1
\end{pmatrix}\begin{matrix} \\ \\ \\ (i) \\ \\ \\ \\ (j) \\ \\ \\ \\ \end{matrix}$$

$$\qquad\qquad\qquad (i) \qquad\qquad\qquad\qquad (j)$$

为 Givens 矩阵（或初等旋转矩阵），简记为 $\boldsymbol{T}_{ij}=\boldsymbol{T}_{ij}(c,\ s)$，由 Givens 矩阵所确定的线性变换叫做 Givens 变换（或初等旋转变换）.

容易验证，当 $c^2+s^2=1$ 时，存在角度 θ，使得 $c=\cos\theta$，$s=\sin\theta$.

Givens 矩阵具有以下性质：

（1）Givens 矩阵是正交矩阵，并且

$$[\boldsymbol{T}_{ij}(c,s)]^{-1}=[\boldsymbol{T}_{ij}(c,s)]^{\mathrm{T}}=\boldsymbol{T}_{ij}(c,-s),\det[\boldsymbol{T}_{ij}(c,s)]=1$$

（2）若 $\boldsymbol{x}=(a_1,\ a_2,\ \cdots,\ a_n)^{\mathrm{T}}$，$\boldsymbol{y}=\boldsymbol{T}_{ij}\boldsymbol{x}=(b_1,\ b_2,\ \cdots,\ b_n)^{\mathrm{T}}$，则有

$$\begin{cases}b_i=ca_i+sa_j\\ b_j=-sa_i+ca_j\\ b_k=a_k \qquad (k\neq i,j)\end{cases}$$

所以，当 $a_i^2+a_j^2\neq 0$ 时，取

$$c=\frac{a_i}{\sqrt{a_i^2+a_j^2}},s=\frac{a_j}{\sqrt{a_i^2+a_j^2}}$$

就可以使 $b_i=\sqrt{a_i^2+a_j^2}>0$，$b_j=0$.

定理 7.2.1 设 $\boldsymbol{x}=(a_1,\ a_2,\ \cdots,\ a_n)^{\mathrm{T}}\neq\boldsymbol{0}$，则存在有限个 Givens 矩阵之积，记作 $\boldsymbol{T}$，使得 $\boldsymbol{T}\boldsymbol{x}=|\boldsymbol{x}|\boldsymbol{e}_1$.

证 如果 $a_1\neq 0$，则对 $\boldsymbol{x}$ 构造 Givens 矩阵 $\boldsymbol{T}_{12}(c,\ s)$

$$c=\frac{a_1}{\sqrt{a_1^2+a_2^2}},s=\frac{a_2}{\sqrt{a_1^2+a_2^2}}$$

$$\boldsymbol{T}_{12}\boldsymbol{x}=(\sqrt{a_1^2+a_2^2},0,a_3,\cdots,a_n)^{\mathrm{T}}$$

再对 $\boldsymbol{T}_{12}\boldsymbol{x}$ 构造 Givens 矩阵 $\boldsymbol{T}_{13}(c,\ s)$

$$c=\frac{\sqrt{a_1^2+a_2^2}}{\sqrt{a_1^2+a_2^2+a_3^2}},s=\frac{a_3}{\sqrt{a_1^2+a_2^2+a_3^2}}$$

$$\boldsymbol{T}_{13}(\boldsymbol{T}_{12}\boldsymbol{x})=(\sqrt{a_1^2+a_2^2+a_3^2},0,0,a_4,\cdots,a_n)^{\mathrm{T}}$$

依此类推，对 $\boldsymbol{T}_{1,n-1}\cdots\boldsymbol{T}_{12}\boldsymbol{x}$ 构造 Givens 矩阵 $\boldsymbol{T}_{1n}(c,\ s)$

$$c=\frac{\sqrt{a_1^2+\cdots+a_{n-1}^2}}{\sqrt{a_1^2+\cdots+a_{n-1}^2+a_n^2}},s=\frac{a_n}{\sqrt{a_1^2+\cdots+a_{n-1}^2+a_n^2}}$$

$$\boldsymbol{T}_{1n}(\boldsymbol{T}_{1,n-1}\cdots\boldsymbol{T}_{12}\boldsymbol{x})=(\sqrt{a_1^2+\cdots+a_{n-1}^2+a_n^2},0,\cdots,0)^{\mathrm{T}}$$

令 $\boldsymbol{T}=\boldsymbol{T}_{1n}\boldsymbol{T}_{1,n-1}\cdots\boldsymbol{T}_{12}$，则有 $\boldsymbol{T}\boldsymbol{x}=|\boldsymbol{x}|\boldsymbol{e}_1$.

如果 $a_1=0$，考虑 $a_1=\cdots=a_{k-1}=0$，$a_k\neq 0$，$(1<k\leqslant n)$ 的情形，此时，$|\boldsymbol{x}|=\sqrt{a_k^2+\cdots+a_n^2}$，由 $\boldsymbol{T}_{1k}$ 开始施行以上的步骤，结论成立.

推论 7.2.1 设非零列向量 $\boldsymbol{x}\in\mathbf{R}^n$ 及单位列向量 $\boldsymbol{z}\in\mathbf{R}^n$，则存在有限个 Givens 矩阵之积，记作 $\boldsymbol{T}$，使得 $\boldsymbol{T}\boldsymbol{x}=|\boldsymbol{x}|\boldsymbol{z}$.

2. Householder 矩阵与 Householder 变换

在二维平面中，将向量 $\boldsymbol{x}$ 映射为关于 Ox 轴对称的向量 $\boldsymbol{y}$ 的变换，叫做关于 Ox 轴的镜像变换. 设 $\boldsymbol{x}=(a_1,\ a_2)^{\mathrm{T}}$，通过镜像变换有

$$\boldsymbol{y}=(a_1,-a_2)^{\mathrm{T}}=\begin{pmatrix}1&0\\0&-1\end{pmatrix}\begin{pmatrix}a_1\\a_2\end{pmatrix}=(\boldsymbol{I}-2\boldsymbol{e}_2\boldsymbol{e}_2^{\mathrm{T}})\boldsymbol{x}=\boldsymbol{H}\boldsymbol{x}$$

其中 $e_2=(0, 1)^T$，H 是正交矩阵，$\det H=-1$.

一般的，对于 n 维欧氏空间而言，将向量 x 映射为关于“与单位向量 u 正交的 $n-1$ 维子空间”对称的向量 y 的镜像变换定义如下.

定义 7.2.2 设单位向量 $u\in \mathbf{R}^n$，称 $H=E-2uu^T$ 为 Householder 矩阵（或初等反射矩阵），由 Householder 矩阵所确定的线性变换叫做 Householder 变换（或初等反射变换）.

Householder 矩阵具有以下性质：

(1) Householder 矩阵是对称矩阵，即 $H=H^T$；

(2) Householder 矩阵是正交矩阵，即 $H^TH=E$；

(3) Householder 矩阵是对合矩阵，即 $H^2=E$；

(4) Householder 矩阵是自逆矩阵，即 $H=H^{-1}$；

(5) $\det H=-1$.

定理 7.2.2 任意给定非零列向量 $x\in \mathbf{R}^n\,(n>1)$ 及单位列向量 $z\in \mathbf{R}^n$，则存在 Householder 矩阵 H，使得 $Hx=|x|z$.

证 当 $x=|x|z$ 时，取单位列向量 u 满足 $u^Tx=0$，则

$$Hx=(E-2uu^T)x=x-2u(u^Tx)=x=|x|z,$$

当 $x\neq |x|z$ 时，取 $u=\dfrac{x-|x|z}{|x-|x|z|}$，则

$$\begin{aligned}Hx&=\left[E-2\frac{(x-|x|z)(x-|x|z)^T}{|x-|x|z|^2}\right]x\\&=x-2(x-|x|z,x)\frac{x-|x|z}{|x-|x|z|^2}\\&=x-(x-|x|z)=|x|z\end{aligned}$$

此处应用了等式 $|x-|x|z|^2=2(x-|x|z,x)$.

定理 7.2.3 Givens 矩阵是两个 Householder 矩阵的乘积，即 $T_{ij}=H_vH_u$. 其中 H_u，H_v 分别为在两个单位向量 u，v 下的 Householder 矩阵.

证

$$T_{ij}=\begin{pmatrix}1&&&&&&&&&&\\&\ddots&&&&&&&&&\\&&1&&&&&&&&\\&&&\cos\theta&&&&\sin\theta&&&\\&&&&1&&&&&&\\&&&&&\ddots&&&&&\\&&&&&&1&&&&\\&&&-\sin\theta&&&&\cos\theta&&&\\&&&&&&&&1&&\\&&&&&&&&&\ddots&\\&&&&&&&&&&1\end{pmatrix}\begin{matrix}\\\\\\(i)\\\\\\\\(j)\\\\\\\\\end{matrix}$$

$$(i)\qquad\qquad(j)$$

取

$$\boldsymbol{u}=\left(0,\quad \cdots,\quad 0\quad \sin\frac{\theta}{4},\quad 0,\quad \cdots,\quad 0,\quad \cos\frac{\theta}{4},\quad 0,\quad \cdots,\quad 0\right)^{\mathrm{T}}$$

则

$$\boldsymbol{H}_u=\boldsymbol{I}-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$$

$$=\begin{pmatrix}1&&&&&&&&&\\&\ddots&&&&&&&&\\&&1&&&&&&&\\&&&\cos\frac{\theta}{2}&&&-\sin\frac{\theta}{2}&&&\\&&&&1&&&&&\\&&&&&\ddots&&&&\\&&&&&&1&&&\\&&&-\sin\frac{\theta}{2}&&&-\cos\frac{\theta}{2}&&&\\&&&&&&&1&&\\&&&&&&&&\ddots&\\&&&&&&&&&1\end{pmatrix}\begin{matrix}\\\\\\(i)\\\\\\\\(j)\\\\\\\\\end{matrix}$$

$$\qquad\qquad(i)\qquad\qquad\qquad(j)$$

取 $\boldsymbol{v}=\left(0,\quad \cdots,\quad 0\quad \sin\frac{3\theta}{4},\quad 0,\quad \cdots,\quad 0,\quad \cos\frac{3\theta}{4},\quad 0,\quad \cdots,\quad 0\right)^{\mathrm{T}}$

则

$$\boldsymbol{H}_v=\boldsymbol{I}-2\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}$$

$$=\begin{pmatrix}1&&&&&&&&&\\&\ddots&&&&&&&&\\&&1&&&&&&&\\&&&\cos\frac{3\theta}{2}&&&-\sin\frac{3\theta}{2}&&&\\&&&&1&&&&&\\&&&&&\ddots&&&&\\&&&&&&1&&&\\&&&-\sin\frac{3\theta}{2}&&&-\cos\frac{3\theta}{2}&&&\\&&&&&&&1&&\\&&&&&&&&\ddots&\\&&&&&&&&&1\end{pmatrix}\begin{matrix}\\\\\\(i)\\\\\\\\(j)\\\\\\\\\end{matrix}$$

$$\qquad\qquad(i)\qquad\qquad\qquad(j)$$

直接计算可得

$$\boldsymbol{T}_{ij}=\boldsymbol{H}_v\boldsymbol{H}_u$$

但是 Householder 矩阵不能由若干个 Givens 矩阵的乘积表示，因为 $\det\boldsymbol{H}=-1$，而 $\det\boldsymbol{T}_{ij}=1$.

3. 矩阵的 QR 分解

定义 7.2.3　设 n 阶复矩阵 $\boldsymbol{A}$ 能够分解为一个 n 阶酉矩阵 $\boldsymbol{Q}$ 和一个 n 阶非奇异上三角矩阵 $\boldsymbol{R}$ 之积（或 n 阶实矩阵 $\boldsymbol{A}$ 能够分解为一个 n 阶正交矩阵 $\boldsymbol{Q}$ 和一个 n 阶非奇异上三角矩阵 $\boldsymbol{R}$ 之积），即 $\boldsymbol{A}=\boldsymbol{QR}$，则称之为矩阵 $\boldsymbol{A}$ 的 QR 分解.

定理 7.2.4　设 $\boldsymbol{A}$ 是任意 n 阶非奇异复矩阵，则矩阵 $\boldsymbol{A}$ 有 QR 分解.

证法一：将矩阵 $\boldsymbol{A}$ 按列分块为 $\boldsymbol{A}=(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n)$，由定理 7.2.2 可知，存在着 n 阶 Householder 矩阵 $\boldsymbol{H}_1$，使得 $\boldsymbol{H}_1\boldsymbol{a}_1=\alpha_1\boldsymbol{e}_1$，因此 $\boldsymbol{H}_1\boldsymbol{A}=(\boldsymbol{H}_1\boldsymbol{a}_1, \boldsymbol{H}_1\boldsymbol{a}_2, \cdots, \boldsymbol{H}_1\boldsymbol{a}_n)$
$=\left(\begin{array}{c|ccc}\alpha_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \boldsymbol{B}_{n-1} & \\ 0 & & & \end{array}\right)$，其中 $\boldsymbol{B}_{n-1}$ 是 $n-1$ 阶复矩阵，再将 $\boldsymbol{B}_{n-1}$ 按列分块为 $\boldsymbol{B}_{n-1}$ $=(\boldsymbol{b}_2, \cdots, \boldsymbol{b}_n)$，则存在着 $n-1$ 阶 Householder 矩阵 $\widetilde{\boldsymbol{H}}_2$，使得 $\widetilde{\boldsymbol{H}}_2\boldsymbol{b}_2=\alpha_2\tilde{\boldsymbol{e}}_1$，（$\tilde{\boldsymbol{e}}_1=(1, 0, \cdots, 0)^{\mathrm{T}}\in \mathbf{C}^{n-1}$），记 $\boldsymbol{H}_2=\begin{pmatrix}1 & \boldsymbol{0}^{\mathrm{T}} \\ 0 & \widetilde{\boldsymbol{H}}_2^{\mathrm{T}}\end{pmatrix}$，则 $\boldsymbol{H}_2$ 是 Householder 矩阵，并且 $\boldsymbol{H}_2$（$\boldsymbol{H}_1\boldsymbol{A}$）$=$
$\left(\begin{array}{c|ccc}\alpha_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \widetilde{\boldsymbol{H}}_2\boldsymbol{B}_{n-1} & \\ 0 & & & \end{array}\right)=\left(\begin{array}{cc|ccc}\alpha_1 & * & * & \cdots & * \\ 0 & \alpha_2 & * & \cdots & * \\ \hline 0 & 0 & & \mathbf{C}_{n-2} & \end{array}\right)$，其中 $\mathbf{C}_{n-2}$ 是 $n-2$ 阶复矩阵，依此类推，

到第 $n-1$ 步有 $\boldsymbol{H}_{n-1}\cdots\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A}=\begin{pmatrix}\alpha_1 & & * \\ & \ddots & \\ & & \alpha_n\end{pmatrix}=\boldsymbol{R}$，其中的 $\boldsymbol{H}_k(k=1, 2, \cdots, n-1)$ 都是 n 阶 Householder 矩阵. 由于 $\boldsymbol{H}_k$ 的自逆性，所以有 $\boldsymbol{A}=\boldsymbol{H}_1\boldsymbol{H}_2\cdots\boldsymbol{H}_{n-1}\boldsymbol{R}=\boldsymbol{QR}$，此处 $\boldsymbol{Q}=\boldsymbol{H}_1\boldsymbol{H}_2\cdots\boldsymbol{H}_{n-1}$ 是酉矩阵，$\boldsymbol{R}$ 是上三角矩阵.

证法二：将矩阵 $\boldsymbol{A}$ 按列分块为 $\boldsymbol{A}=(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n)$，由推论 7.2.1 可知，存在着 n 阶 Givens 矩阵 $\boldsymbol{T}_{12}, \cdots, \boldsymbol{T}_{1n}$，使得

$$\boldsymbol{T}_{1n}\cdots\boldsymbol{T}_{12}\boldsymbol{a}_1=\|\boldsymbol{a}\|_2\boldsymbol{e}_1$$

因此

$$\boldsymbol{T}_{1n}\cdots\boldsymbol{T}_{12}\boldsymbol{A}=\left(\begin{array}{c|ccc}\|\boldsymbol{a}_1\|_2 & * & \cdots & * \\ \hline 0 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_n\end{array}\right),(\boldsymbol{b}_k\in\mathbf{C}^{n-1},k=1,2,\cdots,n)$$

对于其第 2 列，存在 n 阶 Givens 矩阵 $\boldsymbol{T}_{23}, \cdots, \boldsymbol{T}_{2n}$，使得

$$\boldsymbol{T}_{2n}\cdots\boldsymbol{T}_{23}\begin{pmatrix}* \\ \boldsymbol{b}_2\end{pmatrix}=(*,\|\boldsymbol{b}_2\|_2,0,\cdots,0)^{\mathrm{T}}$$

所以

$$\boldsymbol{T}_{2n}\cdots\boldsymbol{T}_{23}\boldsymbol{T}_{1n}\cdots\boldsymbol{T}_{12}\boldsymbol{A}=\left(\begin{array}{cc|ccc}\|\boldsymbol{a}_1\|_2 & * & * & \cdots & * \\ 0 & \|\boldsymbol{b}_2\|_2 & * & \cdots & * \\ \hline 0 & 0 & \boldsymbol{c}_3 & \cdots & \boldsymbol{c}_n\end{array}\right)$$

$(\boldsymbol{c}_k\in\mathbf{C}^{n-2}, k=3, \cdots, n)$

依此类推，最后得到

$$\boldsymbol{T}_{n-1,n}\cdots\boldsymbol{T}_{2n}\cdots\boldsymbol{T}_{23}\boldsymbol{T}_{1n}\cdots\boldsymbol{T}_{12}\boldsymbol{A}=\boldsymbol{R}$$

由此有

$$A=T_{12}^{H}\cdots T_{1n}^{H}T_{23}^{H}\cdots T_{2n}^{H}\cdots T_{n-1,n}^{H}R=QR,$$

其中 $Q=T_{12}^{H}\cdots T_{1n}^{H}T_{23}^{H}\cdots T_{2n}^{H}\cdots T_{n-1,n}^{H}$ 是酉矩阵，R 是上三角矩阵.

这个定理的证明过程给出了用 Householder 变换及用 Givens 变换求矩阵的 QR 分解的方法.

若 A 是可逆矩阵，则有以下的定理.

定理 7.2.5 设 A 是任意 n 阶可逆复矩阵，则矩阵 A 可以唯一的分解为 $A=QR$. 其中 Q 是 n 阶酉矩阵，R 是具有正对角元素的上三角可逆矩阵.

证 将矩阵 A 按列分块为 $A=(a_1, a_2, \cdots, a_n)$，因为矩阵 A 可逆，所以 $a_1, a_2, \cdots, a_n$ 线性无关，用 Schmidt 正交化方法将其正交化

$$\begin{cases} p_1=a_1 \\ p_2=a_2-\lambda_{21}p_1 \\ \vdots \\ p_n=a_n-\lambda_{n1}p_1-\cdots-\lambda_{n,n-1}p_{n-1} \end{cases}$$

其中 $\lambda_{ij}=\dfrac{(a_i, p_j)}{(p_j, p_j)}$. 再将 $p_k(k=1, 2, \cdots, n)$ 单位化，得到

$$q_k=\frac{p_k}{\|p_k\|_2} \quad (k=1,2,\cdots,n)$$

则有

$$\begin{cases} a_1=p_1=\|p_1\|_2q_1 \\ a_2=\lambda_{21}p_1+p_2=\lambda_{21}\|p_1\|_2q_1+\|p_2\|_2q_2 \\ \vdots \\ a_n=\lambda_{n1}p_1+\cdots+\lambda_{n,n-1}p_{n-1}+p_n=\lambda_{n1}\|p_1\|_2q_1+\cdots\lambda_{n,n-1}\|p_{n-1}\|_2q_{n-1}+\|p_n\|_2q_n \end{cases}$$

所以

$$\begin{aligned} A&=(a_1,a_2,\cdots,a_n) \\ &=(q_1,q_2,\cdots,q_n)\begin{pmatrix} \|p_1\|_2 & \lambda_{21}\|p_1\|_2 & \cdots & \lambda_{n1}\|p_1\|_2 \\ & \|p_2\|_2 & \cdots & \lambda_{n2}\|p_2\|_2 \\ & & \ddots & \vdots \\ & & & \|p_n\|_2 \end{pmatrix}=QR \end{aligned}$$

其中 $Q=(q_1, q_2, \cdots, q_n)$ 是 n 阶酉矩阵，R 是具有正对角元素的上三角可逆矩阵.

唯一性：设矩阵 A 有两个 QR 分解 $A=QR=Q_1R_1$，则 $Q=Q_1R_1R^{-1}=Q_1D$，其中 $D=R_1R^{-1}$ 是具有正对角元素的上三角可逆矩阵，由于 $E=Q^HQ=(Q_1D)^H(Q_1D)=D^HD$，则 D 仍为 n 阶酉矩阵，因此 D 是单位矩阵，所以 $Q=Q_1D=Q_1$，$R_1=DR=R$.

在定理 7.2.5 中，若不要求上三角矩阵 R 具有正对角元素，则矩阵 A 的不同的 QR 分解仅在于酉矩阵 Q 的列和上三角矩阵 R 的对应行相差模为 1 的因子. 这个定理的证明过程给出了应用 Schmidt 正交化方法求可逆矩阵 A 的 QR 分解的方法.

总结以上的内容，可以用 Householder 变换，Givens 变换和 Schmidt 正交化方法求矩阵 A 的 QR 分解.

【例 7.2.1】 试求矩阵 $A=\begin{pmatrix}0&3&1\\0&4&-2\\2&1&2\end{pmatrix}$ 的 QR 分解.

解 我们分别应用 Householder 变换，Givens 变换和 Schmidt 正交化方法求矩阵 A 的 QR 分解.

解法一：应用 Householder 变换求 QR 分解.

因为 $a_1=(0,0,2)^{\mathrm{T}}$，取 $\alpha_1=\|a_1\|_2=2$，作单位向量

$$u_1=\frac{a_1-\alpha_1e_1}{\|a_1-\alpha_1e_1\|_2}=\frac{1}{\sqrt{2}}(-1,0,1)^{\mathrm{T}}$$

则

$$H_1=E-2u_1u_1^{\mathrm{T}}=\begin{pmatrix}0&0&1\\0&1&0\\1&0&0\end{pmatrix},H_1A=\begin{pmatrix}2&1&2\\0&4&-2\\0&3&1\end{pmatrix}$$

又因为 $b_2=(4,3)^{\mathrm{T}}$，取 $\alpha_2=\|b_2\|_2=5$，

作单位向量

$$\tilde{u}_2=\frac{b_2-\alpha_2\tilde{e}_1}{\|b_2-\alpha_2e_1^{\mathrm{T}}\|_2}=\frac{1}{\sqrt{10}}(-1,3)^{\mathrm{T}}$$

$$\widetilde{H}_2=E-2\tilde{u}_2u_2^{\mathrm{T}}=\frac{1}{5}\begin{pmatrix}4&3\\3&-4\end{pmatrix}$$

记 $H_2=\begin{pmatrix}1&0^{\mathrm{T}}\\0&\widetilde{H}_2\end{pmatrix}$，则

$$H_2H_1A=\begin{pmatrix}2&1&2\\0&5&-1\\0&0&-2\end{pmatrix}=R$$

所以矩阵 A 的 QR 分解为

$$A=(H_1H_2)R=\begin{pmatrix}0&3/5&-4/5\\0&4/5&3/5\\1&0&0\end{pmatrix}\begin{pmatrix}2&1&2\\0&5&-1\\0&0&-2\end{pmatrix}$$

解法二：应用 Givens 变换求 QR 分解. 取 $c_1=0$，$s_1=1$，则

$$T_{13}=\begin{pmatrix}0&0&1\\0&1&0\\-1&0&0\end{pmatrix},T_{13}A=\begin{pmatrix}2&1&2\\0&4&-2\\0&-3&-1\end{pmatrix}$$

再取 $c_2=\frac{4}{5}$，$s_2=-\frac{3}{5}$，则

$$T_{23}=\begin{pmatrix}1&0&0\\0&4/5&-3/5\\0&3/5&4/5\end{pmatrix},T_{23}T_{13}A=\begin{pmatrix}2&1&2\\0&5&-1\\0&0&-2\end{pmatrix}=R$$

所以矩阵 A 的 QR 分解为

$$A=(T_{13}^{T}T_{23}^{T})R=\begin{pmatrix}0&3/5&-4/5\\0&4/5&3/5\\1&0&0\end{pmatrix}\begin{pmatrix}2&1&2\\0&5&-1\\0&0&-2\end{pmatrix}$$

解法三：应用 Schmidt 正交化方法求 QR 分解. 因为 $a_1=(0,\ 0,\ 2)^T$，$a_2=(3,\ 4,\ 1)^T$，$a_3=(1,\ -2,\ 2)^T$，所以 a_1，a_2，a_3 线性无关，应用 Schmidt 正交化得到

$$p_1=a_1=(0,0,2)^T$$

$$p_2=a_2-\frac{1}{2}p_1=(3,4,0)^T$$

$$p_3=a_3-p_1+\frac{1}{5}p_2=\left(\frac{8}{5},-\frac{6}{5},0\right)^T$$

单位化得到

$$q_1=\frac{1}{2}p_1=(0,0,1)^T$$

$$q_2=\frac{1}{5}p_2=\left(\frac{3}{5},\frac{4}{5},0\right)^T$$

$$q_3=\frac{1}{2}p_3=\left(\frac{4}{5},-\frac{3}{5},0\right)^T$$

因此

$$a_1=p_1=2q_1$$

$$a_2=\frac{1}{2}p_1+p_2=q_1+5q_2$$

$$a_3=p_1-\frac{1}{5}p_2+p_3=2q_1-q_2+2q_3$$

所以矩阵 A 的 QR 分解为

$$A=\begin{pmatrix}0&3/5&-4/5\\0&4/5&3/5\\1&0&0\end{pmatrix}\begin{pmatrix}2&1&2\\0&5&-1\\0&0&2\end{pmatrix}$$

4. QR 分解的 MATLAB 计算实现以及在解方程中的应用

对矩阵 X 进行 QR 分解，就是把 X 分解为一个正交矩阵 Q 和一个上三角矩阵 R 的乘积形式. QR 分解只能对方阵进行. MATLAB 的函数 qr 可用于对矩阵进行 QR 分解，其调用格式为

$$[Q,R]=\mathrm{qr}(X)$$

其产生一个一个正交矩阵 Q 和一个上三角矩阵 R，使之满足 $X=QR$. 以及

$$[Q,R,E]=\mathrm{qr}(X)$$

其产生一个一个正交矩阵 Q、一个上三角矩阵 R 以及一个置换矩阵 E，使之满足 $XE=QR$. 实现 QR 分解后，线性方程组 $Ax=b$ 的解 $x=R\backslash(Q\backslash b)$ 或 $x=E(R\backslash(Q\backslash b)$.

【例 7.2.2】 解如下线性方程组

$$\begin{cases}2x_1+x_2-5x_3+x_4=13\\x_1-5x_2+7x_4=-9\\2x_2+x_3-x_4=6\\x_1+6x_2-x_3-4x_4=0\end{cases}$$

用 QR 分解求解以上的线性方程组．命令如下．

```
A=[2,1,-5,1;1,-5,0,7;0,2,1,-1;1,6,-1,-4];
b=[13,-9,6,0]';
[Q,R]=qr(A);
x=R\(Q\b)
```

或采用 QR 分解的第 2 种格式，命令如下．

```
[Q,R,E]=qr(A);
x=E*(R\(Q\b))
```

运行结果为：

```
x=
   -66.5556
    25.6667
   -18.7778
    26.5556
```

§7.3　矩阵的满秩分解

本节讨论一个 $m\times n$ 复矩阵 $\boldsymbol{A}$ 可以分解为两个与 $\boldsymbol{A}$ 的秩相同的矩阵之积的问题．

定义 7.3.1　设 $m\times n$ 复矩阵 $\boldsymbol{A}$ 的秩为 r，如果存在两个与 $\boldsymbol{A}$ 的秩相同的复矩阵 $\boldsymbol{F}$ 与 $\boldsymbol{G}$，使得 $\boldsymbol{A}=\boldsymbol{FG}$，则称此式为复矩阵 $\boldsymbol{A}$ 的满秩分解．

当 $\boldsymbol{A}$ 是满秩矩阵时（行满秩或列满秩）$\boldsymbol{A}$ 可以分解为单位矩阵与 $\boldsymbol{A}$ 自身的乘积，这个满秩分解叫做平凡分解．

定理 7.3.1　设 $m\times n$ 复矩阵 $\boldsymbol{A}$ 的秩为 $r>0$，则 $\boldsymbol{A}$ 有满秩分解．

证　因为 $\mathrm{rank}\boldsymbol{A}=r>0$，对 $\boldsymbol{A}$ 施行初等行变换，可得到阶梯形矩阵 $\boldsymbol{B}=\begin{pmatrix}\boldsymbol{G}\\0\end{pmatrix}$，其中 $\boldsymbol{G}$ 为 $r\times n$ 矩阵，并且 $\mathrm{rank}\boldsymbol{G}=r>0$．因此存在着有限个 m 阶初等矩阵之积，记作 $\boldsymbol{P}$，有 $\boldsymbol{PA}=\boldsymbol{B}$，或者 $\boldsymbol{A}=\boldsymbol{P}^{-1}\boldsymbol{B}$，将矩阵 $\boldsymbol{P}^{-1}$ 分块为 $\boldsymbol{P}^{-1}=(\boldsymbol{F}\vdots\boldsymbol{S})$，其中 $\boldsymbol{F}$ 为 $m\times r$ 矩阵，$\boldsymbol{S}$ 为 $m\times(n-r)$ 矩阵，并且 $\mathrm{rank}\boldsymbol{F}=r$，$\mathrm{rank}\boldsymbol{S}=n-r$．

则有 $\boldsymbol{A}=\boldsymbol{P}^{-1}\boldsymbol{B}=(\boldsymbol{F}\vdots\boldsymbol{S})\begin{pmatrix}\boldsymbol{G}\\0\end{pmatrix}=\boldsymbol{FG}$，其中 $\boldsymbol{F}$ 是列满秩矩阵，$\boldsymbol{S}$ 是行满秩矩阵．但是，矩阵 $\boldsymbol{A}$ 的满秩分解不唯一．这是因为若取任意一个 r 阶非奇异矩阵 $\boldsymbol{D}$，则有

$$\boldsymbol{A}=\boldsymbol{FG}=(\boldsymbol{FD})(\boldsymbol{D}^{-1}\boldsymbol{G})=\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{G}}.$$

【例 7.3.1】　求矩阵 $\boldsymbol{A}=\begin{pmatrix}-1&0&1&2\\1&2&-1&1\\2&2&-2&-1\end{pmatrix}$ 的满秩分解．

解　对矩阵 $\boldsymbol{A}$ 进行初等行变换

$$(\boldsymbol{A}\vdots\boldsymbol{E})=\begin{pmatrix}-1&0&1&2&1&0&0\\1&2&-1&1&0&1&0\\2&2&-2&-1&0&0&1\end{pmatrix}\to\begin{pmatrix}-1&0&1&2&1&0&0\\0&2&0&3&1&1&0\\0&0&0&0&1&-1&1\end{pmatrix}=\boldsymbol{B}=\begin{pmatrix}\boldsymbol{G}\\0\end{pmatrix}$$

其中 $G=\begin{pmatrix}-1 & 0 & 1 & 2\\ 0 & 2 & 0 & 3\end{pmatrix}$，所以 $B=\begin{bmatrix}-1 & 0 & 1 & 2\\ 0 & 2 & 0 & 3\\ 0 & 0 & 0 & 0\end{bmatrix}$，$P=\begin{bmatrix}1 & 0 & 0\\ 1 & 1 & 0\\ 1 & -1 & 1\end{bmatrix}$. 而 $P^{-1}=\begin{bmatrix}1 & 0 & 0\\ -1 & 1 & 0\\ -2 & 1 & 1\end{bmatrix}=(F\vdots S)$，其中 $F=\begin{bmatrix}1 & 0\\ -1 & 1\\ -2 & 1\end{bmatrix}$.

由此可见，所以有 $A=P^{-1}B=(F\vdots S)\begin{pmatrix}G\\ 0\end{pmatrix}=FG=\begin{bmatrix}1 & 0\\ -1 & 1\\ -2 & 1\end{bmatrix}\begin{pmatrix}-1 & 0 & 1 & 2\\ 0 & 2 & 0 & 3\end{pmatrix}$.

定义 7.3.2 设 $m\times n$ 复矩阵 H 的秩为 $r(r>0)$，并且满足以下条件：

(1) 矩阵 H 的前 r 行中的每一行至少含有一个不为零的元素，并且第一个不为零的元素是 1，而后 $m-r$ 行的元素均为零；

(2) 如果矩阵 H 的第 i 行的第一个不为零的元素 1 在第 j_i 列 $(i=1,2,\cdots,r)$，则 $j_1<j_2<\cdots<j_r$；

(3) 矩阵 H 的 $j_1,j_2,\cdots,j_r$ 列是单位矩阵 E_m 的前 r 列.

则称矩阵 H 为 **Hermite** 标准形（最简型）.

由此定义可见，对于任意一个秩为 r 的 $m\times n$ 复矩阵 A，均可以经过初等行变换将其化为 **Hermite** 标准形 H，而且矩阵 H 的 $j_1,j_2,\cdots,j_r$ 列元素组成的列向量组线性无关.

定义 7.3.3 以 n 阶单位矩阵 E_n 的 n 个列向量 $e_1,e_2,\cdots,e_n$ 为列构成的 n 阶矩阵 $P=(e_{j_1},e_{j_2},\cdots,e_{j_n})$ 叫做置换矩阵. 其中 $j_1,j_2,\cdots,j_n$ 是 $1,2,\cdots,n$ 的一个全排列.

定理 7.3.2 设 $m\times n$ 复矩阵 A 的秩为 $r(r>0)$，矩阵 A 的 Hermite 标准形为 H，则在矩阵 A 的满秩分解 $A=FG$ 中，可以取矩阵 F 为 A 的 $j_1,j_2,\cdots,j_r$ 列构成的 $m\times r$ 列矩阵，G 为 H 的前 r 行构成的 $r\times n$ 列矩阵.

证 设 $A\xrightarrow{\text{行的初等变换}}H$，则存在 m 阶单位矩阵 P，使得

$$PA=H,A=P^{-1}H$$

根据定理 7.3.1，将 P^{-1} 分块为 $(F\vdots S)$，可得满秩分解 $A=FG$，其中 G 为 H 的前 r 行构成的 $r\times n$ 列矩阵.

参照矩阵 A 的 Hermite 标准形为 H，作置换矩阵 $Q=(e_{j_1},e_{j_2},\cdots,e_{j_n})$，则

$$AQ=(a_{j_1},\ \cdots,\ a_{j_r},\ a_{j_{r+1}},\ \cdots,\ a_{j_n})$$

$$HQ=(h_{j_1},\ \cdots,\ h_{j_r},\ h_{j_{r+1}},\ \cdots,\ h_{j_n})=\begin{pmatrix}E_r & B\\ 0 & 0\end{pmatrix}$$

$$AQ=(P^{-1}H)Q=P^{-1}(HQ)=(F\vdots S)\begin{pmatrix}E_r & B\\ 0 & 0\end{pmatrix}=(F\vdots FB),$$

即 F 为 AQ 的前 r 列构成的矩阵，也就是 A 的 $j_1,j_2,\cdots,j_r$ 列构成的 $m\times r$ 列矩阵.

【例 7.3.2】 求矩阵 $A=\begin{bmatrix}-1 & 0 & 1 & 2\\ 1 & 2 & -1 & 1\\ 2 & 2 & -2 & -1\end{bmatrix}$ 的满秩分解.

解　先求出矩阵 $\boldsymbol{A}$ 的 Hermite 标准形 $\boldsymbol{A}=\begin{bmatrix}-1&0&1&2\\1&2&-1&1\\2&2&-2&-1\end{bmatrix}\to\begin{bmatrix}1&0&-1&2\\0&1&0&3/2\\0&0&0&0\end{bmatrix}=$ $\boldsymbol{H}$，$\boldsymbol{H}$ 的第 1 列与第 2 列构成 $\boldsymbol{E}_3$ 的前两列，所以矩阵 $\boldsymbol{F}$ 为 $\boldsymbol{A}$ 的第 1 列与第 2 列构成的 3×2 矩阵，$\boldsymbol{G}$ 为 $\boldsymbol{H}$ 的前 2 行构成的 2×4 矩阵，即 $\boldsymbol{F}=\begin{bmatrix}-1&0\\1&2\\2&2\end{bmatrix}$，$\boldsymbol{G}=\begin{pmatrix}1&0&-1&2\\0&1&0&3/2\end{pmatrix}$，所以

$$\boldsymbol{A}=\boldsymbol{FG}=\begin{bmatrix}-1&0\\1&2\\2&2\end{bmatrix}\begin{pmatrix}1&0&-1&2\\0&1&0&3/2\end{pmatrix}.$$

对比例 7.3.1，可以看出矩阵 $\boldsymbol{A}$ 的满秩分解不唯一．实际上，如果 $\boldsymbol{A}=\boldsymbol{FG}$ 是 $\boldsymbol{A}$ 的满秩分解，rank（$\boldsymbol{A}$）$=r$，$\boldsymbol{P}$ 是 r 阶可逆矩阵，取 $\boldsymbol{F}_1=\boldsymbol{FP}$，$\boldsymbol{G}_1=\boldsymbol{P}^{-1}\boldsymbol{G}$，则 $\boldsymbol{A}=\boldsymbol{F}_1\boldsymbol{G}_1$ 也是 $\boldsymbol{A}$ 的满秩分解．两个不同的满秩分解有何关系呢？

定理 7.3.3　设 $\boldsymbol{A}=\boldsymbol{FG}=\boldsymbol{F}_1\boldsymbol{G}_1$ 都是 $\boldsymbol{A}$ 的满秩分解，rank（$\boldsymbol{A}$）$=r$，则存在 r 阶可逆矩阵 $\boldsymbol{P}$，使得

$$\boldsymbol{F}_1=\boldsymbol{FP},\boldsymbol{G}_1=\boldsymbol{P}^{-1}\boldsymbol{G}$$

§7.4　矩阵的奇异值分解

矩阵的 Jordan 标准形有两个局限：其一，是只有方阵才能求其 Jordan 标准形；其二，Jordan 标准形毕竟不如对角矩阵来得方便．本节讨论的矩阵奇异值分解，将克服这些局限性．

定理 7.4.1　如果 $\boldsymbol{A}$ 为 n 阶复矩阵，则有：

（1）矩阵 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$，$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的特征值都是非负实数；

（2）矩阵 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 与 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的非零特征值都相同．

证　（1）设 $\boldsymbol{\alpha}\in\mathbf{C}^n$ 为 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的特征值 λ 所对应的特征向量，则 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是 Hermite 矩阵，所以 λ 是实数；并且 $0\leqslant(\boldsymbol{A\alpha},\boldsymbol{A\alpha})=(\boldsymbol{\alpha},\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A\alpha})=(\boldsymbol{\alpha},\lambda\boldsymbol{\alpha})=\lambda(\boldsymbol{\alpha},\boldsymbol{\alpha})$，因为 $\boldsymbol{\alpha}\neq0$，所以 $\lambda\geqslant0$.

同理可证，$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的特征值也是非负实数．

（2）将 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的特征值按顺序记为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_r>\lambda_{r+1}=\lambda_{r+2}=\cdots=\lambda_n=0$，设 $\boldsymbol{\alpha}_i\in\mathbf{C}^n$ $(i=1,2,\cdots,r)$ 为 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的非零特征值 $\lambda_i(i=1,2,\cdots,r)$ 所对应的特征向量，则由 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A\alpha}_i=\lambda_i\boldsymbol{\alpha}_i(i=1,2,\cdots,r)$，有（$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$）$\boldsymbol{A\alpha}_i=\lambda_i\boldsymbol{A\alpha}_i(i=1,2,\cdots,r)$，因为 $\boldsymbol{A\alpha}_i$ 是非零向量，所以 λ_i 也是 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的非零特征值．

同理可证，$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的非零特征值也是 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的非零特征值．

以下证明 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 与 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的非零特征值完全相同，这只要证明 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 与 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的非零特征值的代数重数相同即可．

设 $\boldsymbol{y}_1$，$\boldsymbol{y}_2$，…，$\boldsymbol{y}_p$ 为 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 对应于非零特征值 λ 的线性无关的特征向量，因为 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是 Hermite 矩阵，也就是说 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 既是正规矩阵，也是单纯矩阵．所以 p 就是非零特征值 λ 的代数重数．而 $\boldsymbol{Ay}_i$ 也是 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 对应于非零特征值 λ_i 的特征向量$(i=1,2,\cdots,p)$．而这些向量线性无关，这是因为若 $\boldsymbol{A}$（$\boldsymbol{y}_1$，$\boldsymbol{y}_2$，…，$\boldsymbol{y}_p$）$\boldsymbol{K}=k_1\boldsymbol{Ay}_1+k_2\boldsymbol{Ay}_2+\cdots+k_p\boldsymbol{Ay}_p=\mathbf{0}$，则 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$

(y_1, y_2, …, y_p) $K=0$, 即 λ (y_1, y_2, …, y_p) $K=0$. 由于 $\lambda\neq 0$, 所以 (y_1, y_2, …, y_p) $K=0$, 但 y_1, y_2, …, y_p 线性无关, 所以 $K=0$. 因此, λ 也是 AA^H 的 p 重非零特征值.

对于 Hermite 矩阵 A, 存在酉矩阵 U, 使得 $U^HAU=\text{diag}$ (λ_1, …, λ_r, λ_{r+1}…, λ_n), 其中 λ_1, …, λ_r, λ_{r+1}…, λ_n 是 A 的特征值. 假定 λ_1, …, λ_r 是 A 的非零特征值, 将 U 分块成

$$U=(U_1U_2),U_1\in C^{n\times r}U_2\in C^{n\times(n-r)}$$

则

$$A=U_1\text{diag}(\lambda_1,\cdots,\lambda_r)U_1^H$$

称上式为 Hermite 矩阵 A 的谱分解.

定义 7.4.1 设 A 是秩为 r 的 $m\times n$ 复矩阵, A^HA 的特征值为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_r>\lambda_{r+1}=\lambda_{r+2}=\cdots=\lambda_n=0$ 则 $\sigma_i=\sqrt{\lambda_i}(i=1, 2, \cdots, r)$ 叫做矩阵 A 的正奇异值.

定义 7.4.2 设 A、B 是 $m\times n$ 复矩阵, 若存在 m 阶酉矩阵 U, n 阶酉矩阵 V, 使得 $A=UBV$, 则称矩阵 A 与 B 酉等价.

定理 7.4.2 设 A、B 是 $m\times n$ 复矩阵, 若 A 与 B 酉等价, 则它们有相同的正奇异值.

证 因为 A 与 B 酉等价, 即存在 m 阶酉矩阵 U 与 n 阶酉矩阵 V, 使得 $A=UBV$, 有酉矩阵的性质可知 $U^H=U^{-1}$, $V^H=V^{-1}$, 所以 $A^H=V^HB^HU^H=V^{-1}B^HU^{-1}$, 则 $AA^H=UBVV^{-1}B^HU^{-1}=U$ (BB^H) U^{-1}, 即 AA^H 与 BB^H 酉相似, 所以, AA^H 与 BB^H 有相同的特征值, 即有相同的正奇异值.

定理 7.4.3 (奇异值分解定理) 设 A 是秩为 r 的 $m\times n$ 复矩阵, 则存在 m 阶酉矩阵 U, n 阶酉矩阵 V, 使得 $A=U\begin{pmatrix}\Sigma & 0\\ 0 & 0\end{pmatrix}V^H$. 其中 $\Sigma=\text{diag}$ (δ_1, δ_2, …, δ_r), $|\delta_i|=\sigma_i$ ($i=1$, 2, …, r), $\delta_i\in C$, $\sigma_i(i=1, 2, \cdots, r)$ 是矩阵 A 的正奇异值.

证 记 AA^H 的特征值为

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_r>,\lambda_{r+1}=\lambda_{r+2}=\cdots=\lambda_m=0$$

则存在 m 阶酉矩阵 U, 使得

$$U^H(AA^H)U=\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}=\begin{pmatrix}\Sigma^2 & 0\\ 0 & 0\end{pmatrix}$$

将 U 分块为

$$U=(U_1\,U_2),U_1\in C^{m\times r},U_2\in C^{m\times(m-r)}$$

则有

$$(AA^H)U=(AA^HU_1\,AA^HU_2)=(U_1\,U_2)\begin{pmatrix}\Sigma^2 & 0\\ 0 & 0\end{pmatrix}=(U_1\Sigma^2\quad 0)$$

故

$$U_1^HAA^HU_1=U_1^HU_1\Sigma^2=\Sigma^2,\ U_2^HAA^HU_2=0$$

由此可得 $A^HU_2=0$. 令 $V_1=A^HU_1$ (Σ^{-1}), 则 $V_1^HV_1=E_r$, 即 $V_1=$ (v_1, …, v_r) 的 r 列是两两正交的单位向量. 添加 $n-r$ 单位向量 v_{r+1}, …, v_n, 使 v_1, …, v_r, v_{r+1}, …, v_n 成为 C^n 的标准正交基, 则 $V=$ (v_1, …, v_r, v_{r+1}, …, v_n) 是 n 阶酉矩阵. 记 $V_2=$ (v_{r+1}, …, v_n), 则 $U_2^HU_1=0$.

$$V^H A^H U = V^H (A^H U_1 \quad A^H U_2) = \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} (V_1 \Sigma \quad 0) = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}$$

故

$$A = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^H$$

由定理 7.4.3 有，$A^H A = V \begin{pmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{pmatrix} V^H$，因此 v_j 是 $A^H A$ 的对应于特征值 λ_j 的单位特征向量. 可以验证，$U_1 = AV_1 \Sigma^{-1}$.

由于

$$A = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^H = U_1 \Sigma V_1^H$$

我们也称 $U_1 \Sigma V_1^H$ 为 A 的奇异值分解.

由上述证明可得出，求 A 的奇异值分解的步骤如下：

(1) 求 $A^H A$（或 AA^H）的非零特征值 λ_1，…，λ_r；

(2) 求 AA^H 的对应于特征值 λ_1，…，λ_r 的单位特征向量 u_1，…，u_r；

(3) 求 $A^H A$ 的对应于特征值 λ_1，…，λ_r 的单位特征向量 v_1，…，v_r；

(4) $A = U_1 \Sigma V_1^H$ 为 A 的奇异值分解，其中 $U_1 = (u_1, \cdots, u_r)$，$V_1 = (v_1, \cdots, v_r)$，$\Sigma = \mathrm{diag}(\delta_1, \delta_2, \cdots, \delta_r)$.

【例 7.4.1】 求矩阵 $A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ 的奇异值分解.

解 因为 $AA^H = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ 显然矩阵 AA^H 的特征值为 $\lambda_1 = 5$，$\lambda_2 = \lambda_3 = 0$. 所以，矩阵 A 的正奇异值为 $\sqrt{5}$. 而对应的单位正交特征向量分别为 $\alpha_1 = (1, 0, 0)^T$，$\alpha_2 = (0, 1, 0)^T$，$\alpha_3 = (0, 0, 1)^T$，则 $U = (U_1 \vdots U_2)$，$U_1 = (\alpha_1)$，$U_2 = (\alpha_2, \alpha_3)$，$V_1 = A^H U_1 (\Sigma^{-1}) = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \frac{1}{\sqrt{5}} = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)^T$，取 $V_2 = \left(-\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right)^T$，所以，$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$.

MATLAB 奇异值分解，只要在 matlab 命令行输入 svd（X）就行了.

```
>>X=[2 3; 4 5; 6 8; 4 9];
```

方法一：

```
>>[U S V]=svd(X)
```

方法二：

```
>>[U S V]=svd(X,0)
```

最小奇异值作为静态电压稳定指标已广泛应用于电力系统的稳定分析之中. 在对电力系统进行静态分析时，极坐标形式的潮流方程为

$$\begin{cases} P_{is} = U_i \sum\limits_{j\in i} U_j(G_{ij}\cos\theta_{ij} + B_{ij}\sin\theta_{ij}) \\ Q_{is} = U_i \sum\limits_{j\in i} U_j(G_{ij}\sin\theta_{ij} - B_{ij}\cos\theta_{ij}) \end{cases} \tag{7.4.1}$$

式中：i、j 为节点编号，$j\in i$ 表示 j 必须与 i 直接相连并包含 $j=i$ 的情况；U_i 和 U_j 分别为节点 i 和 j 的电压幅值；G_{ij} 和 B_{ij} 分别为节点导纳矩阵元素的实部与虚部；$\theta_{ij}=\theta_i-\theta_j$ 为节点 i、j 电压相角的差值；P_{is}、Q_{is} 分别为给定节点的有功及无功功率. 在运行点附近将系统的潮流方程式（7.4.1）线性化，则可得到修正方程为

$$\begin{pmatrix} \Delta\boldsymbol{P} \\ \Delta\boldsymbol{Q} \end{pmatrix} = \begin{bmatrix} \boldsymbol{J}_{p\theta} & \boldsymbol{J}_{pv} \\ \boldsymbol{J}_{q\theta} & \boldsymbol{J}_{qv} \end{bmatrix} \begin{pmatrix} \Delta\boldsymbol{\theta} \\ \Delta\boldsymbol{U}/\boldsymbol{U} \end{pmatrix} \tag{7.4.2}$$

式中：$\Delta\boldsymbol{P}$、$\Delta\boldsymbol{Q}$ 分别为节点有功和无功功率的不平衡量子列向量；$\Delta\boldsymbol{\theta}$、$\Delta\boldsymbol{U}$ 分别为节点电压相角与电压幅值的不平衡量子列向量；$\boldsymbol{J}_{p\theta}$，$\boldsymbol{J}_{pv}$，$\boldsymbol{J}_{q\theta}$，$\boldsymbol{J}_{qv}$ 为分块矩阵，其中 $\boldsymbol{J}_{p\theta}$，$\boldsymbol{J}_{pv}$ 中的元素分别为节点有功功率对节点电压角度 q 和电压幅值 U 的偏导数，$\boldsymbol{J}_{q\theta}$，$\boldsymbol{J}_{qv}$ 的元素分别为节点无功功率对节点电压角度 q 和电压幅值 U 的偏导数.

用矩阵 $\boldsymbol{J}$ 表示式（7.4.2）中的雅可比矩阵，若 $J\in\mathbf{R}^{m\times n}$，对其进行奇异值分解可得，

$$\boldsymbol{A}=\boldsymbol{U}\sum\boldsymbol{V}^{\mathrm{T}}$$

令 $\delta_{\min}$ 为最小奇异值，当 $\delta_{\min}\neq 0$ 时，有

$$\begin{pmatrix} \Delta\boldsymbol{\theta} \\ \Delta\boldsymbol{U}/\boldsymbol{U} \end{pmatrix} = \boldsymbol{V}\sum{}^{-1}\boldsymbol{U}^{\mathrm{T}} \begin{pmatrix} \Delta\boldsymbol{P} \\ \Delta\boldsymbol{Q} \end{pmatrix} \tag{7.4.3}$$

由式（7.4.3）可知，当一个奇异值几乎为零时，任何小的扰动将会造成状态变量的巨大变化，当系统处于弱稳定状态、接近于电压崩溃点时，系统响应完全由最小奇异值和它所对应的奇异向量 $\boldsymbol{u}_{\min}$，$\boldsymbol{v}_{\min}$ 决定. 因此雅可比矩阵的最小奇异值的大小表明了电压稳定的相对临近程度，最小奇异值对应的左、右奇异向量提供了进行稳定分析的重要信息. 令 $\delta_{\min}$ 对应的左、右奇异向量分别为 $\boldsymbol{u}_{\min}$ 和 $\boldsymbol{v}_{\min}$. 负荷节点相对于最小奇异值 $\delta_{\min}$ 的参与因子 p 的大小则表明了节点电压的稳定程度，其值越大，则该节点的稳定性越差，p 的表达式为

$$p = \boldsymbol{v}_{\min,i}\boldsymbol{u}_{\min,i} \tag{7.4.4}$$

式中：$\boldsymbol{u}_{\min,i}$ 为向量 $\boldsymbol{u}$ 中对应节点 i 的数值；$\boldsymbol{v}_{\min,i}$ 为向量 $\boldsymbol{v}$ 中对应节点 i 的数值. 进一步的讨论略.

习　题　7

7.1　已知 $\boldsymbol{A}=\begin{bmatrix} 5 & 2 & -4 & 0 \\ 2 & 1 & -2 & 1 \\ -4 & -2 & 5 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix}$，$\boldsymbol{b}=\begin{bmatrix} -3 \\ 2 \\ 7 \\ 10 \end{bmatrix}$，求矩阵 $\boldsymbol{A}$ 的 $\boldsymbol{LU}$ 分解，并利用 $\boldsymbol{LU}$ 分解求解线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$.

7.2　下述矩阵能否进行 $\boldsymbol{LU}$ 分解？如果能分解，分解式是否唯一？

$$A=\begin{pmatrix}1&2&3\\2&4&1\\4&6&7\end{pmatrix},B=\begin{pmatrix}1&1&1\\2&2&1\\3&3&1\end{pmatrix},C=\begin{pmatrix}1&2&6\\2&5&15\\6&15&46\end{pmatrix}$$

7.3 分别应用 Schmidt 正交化方法、Givens 变换以及 Householder 变换求下列矩阵的 QR 分解.

$$A=\begin{pmatrix}0&1&1\\1&1&0\\1&0&1\end{pmatrix},B=\begin{pmatrix}2&2&1\\0&2&2\\2&1&2\end{pmatrix},C=\begin{pmatrix}0&4&1\\1&1&1\\0&3&2\end{pmatrix}$$

7.4 求矩阵 $A=\begin{pmatrix}1&4&-1&5&6\\2&0&0&4&0\\-1&2&-4&-4&-4\\1&-2&-1&-1&-6\end{pmatrix}$ 的满秩分解.

7.5 求矩阵 $A=\begin{pmatrix}1&0\\0&1\\2&0\end{pmatrix}$ 的奇异值分解.

7.6 若矩阵 $A\in\mathbf{R}^{n\times n}$，$A=A^{\mathrm{T}}$ 并且 $A^2=0$，试证明 $A=0$.

7.7 证明：若矩阵 $A\in\mathbf{R}^{n\times n}$，$A=A^{\mathrm{T}}$，必存在方阵 B，$B=B^{\mathrm{T}}$ 使得 $B^3=A$.

7.8 证明：若矩阵 $A\in\mathbf{R}^{n\times n}$，$A=A^{\mathrm{T}}$，并且 A 的主对角元素都大于零，则 A 至少有一个正特征值.

7.9 设矩阵 A 的最大秩分解为 $A=BC$，证明 $Ax=0$ 的充分且必要条件是 $Cx=0$.

7.10 设 a_1 和 a_2 是 n 阶方阵 A 的最大奇异值和最小奇异值，证明：$\|A\|_2=a^1$；当 A 可逆时，$\|A^{-1}\|_2=\dfrac{1}{a_2}$.

第八章 广义逆矩阵

在线性代数中，学习了逆矩阵的概念. 我们知道，对于方阵A，若$|A|\neq 0$，则存在唯一的方阵B使得

$$AB = BA = E$$

并称B是A的逆矩阵，记为A^{-1}. 当A不是方阵或$|A|=0$时，上述逆不存在. 但是在实际问题中，遇到的矩阵不一定是方阵，即使是方阵也不一定是非奇异的. 因此需要将逆矩阵的概念进一步推广.

早在1920年，E. H. Moore就提出了广义逆矩阵的概念. 但在其后的30年中，由于不知其用途，他的理论几乎未被注意. 直到1955年，R. Penrose以更明确的形式给出了与E. H. Moore等价的广义逆矩阵的定义，广义逆矩阵的研究才进入了一个新阶段. 广义逆矩阵由于在数据分析、多元分析、信号处理、系统理论、现代控制理论和网络理论等许多领域中有着重要的应用，使得这一学科得到迅速发展，并成为矩阵论的一个重要分支.

本章中，我们重点介绍应用最多的五类广义逆矩阵的概念、性质和计算方法，其他类型的广义逆矩阵只做简单的介绍.

§8.1 广义逆矩阵及其分类

定义 8.1.1 设$A\in C^{m\times n}$，若$G\in C^{n\times m}$满足如下 Penrose-Moore 方程

(1) $AGA=A$;

(2) $GAG=G$;

(3) $(AG)^{H}=AG$;

(4) $(GA)^{H}=GA$.

的全部或一部分，则称G是A的广义逆矩阵.

按照这一定义，广义逆矩阵可以分为满足其中一个方程的广义逆矩阵、满足其中两个方程的广义逆矩阵、满足其中三个方程的广义逆矩阵和满足全部方程的广义逆矩阵，因而共有15类广义逆矩阵.

如果G是满足第i个方程的广义逆矩阵，就记为$A^{(i)}$；如果G是满足第i，j个方程的广义逆矩阵，就记为$A^{(i,j)}$；如果G是满足第i，j，k个方程的广义逆矩阵，就记为$A^{(i,j,k)}$；如果G是满足四个方程的广义逆矩阵，就记为$A^{(1,2,3,4)}$.

下面我们将会看到，除了$A^{(1,2,3,4)}$是唯一确定之外，其余各类广义逆矩阵都不是唯一确定的，每一类广义逆矩阵都包含着一类矩阵，为了表示这种情况，把满足前面所述相应条件的一切广义逆矩阵分别记为

$$A\{i\},A\{i,j\},A\{i,j,k\}$$

在上述十五类广义逆矩阵中，应用较多的是以下五种：

(1) $A\{1\}$，其中任意一个固定广义逆矩阵记为 A^-；

(2) $A\{1,2\}$，其中任意一个固定广义逆矩阵记为 A_r^-；

(3) $A\{1,3\}$，其中任意一个固定广义逆矩阵记为 A_m^-；

(4) $A\{1,4\}$，其中任意一个固定广义逆矩阵记为 A_l^-；

(5) $A^+=A^{\{1,2,3,4\}}$.

因为 A^+ 满足全部四个方程，显然有

$$A^+\in A\{1\},A^+\in A\{1,2\},A^+\in A\{1,3\},A^+\in A\{1,4\}$$

因此 A^+ 在广义逆矩阵中占有十分重要的位置.

§8.2 广义逆矩阵 A^-

广义逆矩阵 A^- 起源于线性方程组

$$Ax=b$$

(其中 A 是 $m\times n$ 矩阵) 的求解问题. 如果方程组 $Ax=b$ 中 A 是一个可逆矩阵，那么它的解可表示为 $x=A^{-1}b$，而且解是唯一的. 在一般情况下，方程组 $Ax=b$ 有解的充分必要条件：

(1) rank (A,b) =rank (A)；

(2) $b\in R(A)$.

由此自然会想到它的解是否也可用某个 $n\times m$ 矩阵 G 表示为 $x=Gb$ 呢？如果它的解能表示为 $x=Gb$，那么矩阵 G 如何求，以及 G 具有什么性质？下面就来回答这两个问题.

定理 8.2.1 对任意的 $b\in R(A)$，存在矩阵 G 使得 $x=Gb$ 都是方程 $Ax=b$ 的解的充分必要条件是 G 满足 $AGA=A$.

证 先证必要性. 设对任意的 $b\in R(A)$，存在矩阵 G 使得 $x=Gb$ 都是方程 $Ax=b$ 的解，则对 $b_i=Ae_i$，$e_i=(0,\cdots,0,\underset{(i)}{1},0,\cdots,0)^T$，$x=Gb_i=GAe_i$ 是 $Ax=b_i$ 的一个解，所以

$$AGAe_i=Ae_i,(i=1,2,\cdots,n).$$

$$AGAE=AGA(e_1,e_2,\cdots,e_n)=(AGAe_1,AGAe_2,\cdots,AGAe_n)$$
$$=(Ae_1,Ae_2,\cdots,Ae_n)=AE=A$$

故 $AGA=A$.

再证充分性. 设 G 满足 $AGA=A$，则对任意的 $b\in R(A)$，存在 $z\in C^n$，使得 $b=Az$，因此 $AGAz=Az=b$，即 $x=Gb$ 是方程 $Ax=b$ 的解.

定义 8.2.1 设 $A\in C^{m\times n}$，若矩阵 $G\in C^{n\times m}$ 使得 $AGA=A$，则称 G 是 A 的减号逆，记作 A^-，或记 $G\in A\{1\}$.

引理 8.2.1 设 A 是 $m\times n$ 矩阵，rank $(A)=r$，则存在非奇异矩阵 $P\in C^{m\times m}$，$Q\in C^{n\times n}$ 使得

$$PAQ=\begin{pmatrix}E_r & 0\\ 0 & 0\end{pmatrix}$$

证 由线性代数知识知道，对矩阵 A，可以利用一系列行的初等变换将 A 化为如下最简型

$$B=\begin{pmatrix}1&0&0&\cdots&0&b_{1,r+1}&\cdots&b_{1,n}\\0&1&0&\cdots&0&b_{2,r+1}&\cdots&b_{2,n}\\0&0&1&\cdots&0&b_{3,r+1}&\cdots&b_{3,n}\\\vdots&\vdots&\vdots&\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&\cdots&1&b_{r,r+1}&\cdots&b_{r,n}\\0&0&0&\cdots&0&0&\cdots&0\\\vdots&\vdots&\vdots&\vdots&\vdots&\vdots&\vdots&\vdots\\0&0&0&\cdots&0&0&\cdots&0\end{pmatrix}$$

再对它进行列变换，就可得到矩阵$\begin{pmatrix}E_r&0\\0&0\end{pmatrix}$. 而对矩阵$A$进行行变换，相当于用相应的初等方阵去左乘$A$，而对矩阵$A$进行列变换，相当于用相应的初等方阵去右乘$A$. 因此存在非奇异矩阵$P\in C^{m\times m}$，$Q\in C^{n\times n}$使得

$$PAQ=\begin{pmatrix}E_r&0\\0&0\end{pmatrix}$$

定理 8.2.2 设A是$m\times n$矩阵，rank（A）$=r$，非奇异矩阵$P\in C^{m\times m}$，$Q\in C^{n\times n}$使得

$$PAQ=\begin{pmatrix}E_r&0\\0&0\end{pmatrix}$$

则$G\in C^{n\times m}$使得$AGA=A$的充分必要条件是

$$G=Q\begin{bmatrix}E_r&G_{12}\\G_{21}&G_{22}\end{bmatrix}P$$

其中G_{12}，G_{21}，G_{22}分别是$r\times(m-r)$，$(n-r)\times r$，$(n-r)\times(m-r)$ 的任意矩阵.

证 先证必要性. 由定理所设，存在非奇异矩阵P，Q使得

$$PAQ=\begin{pmatrix}E_r&0\\0&0\end{pmatrix}$$

从而有

$$A=P^{-1}\begin{pmatrix}E_r&0\\0&0\end{pmatrix}Q^{-1}$$

由$AGA=A$得

$$P^{-1}\begin{pmatrix}E_r&0\\0&0\end{pmatrix}Q^{-1}GP^{-1}\begin{pmatrix}E_r&0\\0&0\end{pmatrix}Q^{-1}=P^{-1}\begin{pmatrix}E_r&0\\0&0\end{pmatrix}Q^{-1}$$

$$\begin{pmatrix}E_r&0\\0&0\end{pmatrix}Q^{-1}GP^{-1}\begin{pmatrix}E_r&0\\0&0\end{pmatrix}=\begin{pmatrix}E_r&0\\0&0\end{pmatrix}\tag{8.2.1}$$

设

$$Q^{-1}GP^{-1}=\begin{bmatrix}G_{11}&G_{12}\\G_{21}&G_{22}\end{bmatrix}$$

将它代入（8.2.1）有$G_{11}=E_r$，因此

$$G=Q\begin{bmatrix}E_r&G_{12}\\G_{21}&G_{22}\end{bmatrix}P$$

其中 G_{12}，G_{21}，G_{22} 分别是 $r\times(m-r)$，$(n-r)\times r$，$(n-r)\times(m-r)$ 的任意矩阵.

再证充分性. 由定理所设，存在非奇异矩阵 P，Q 使得

$$PAQ=\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$$

而

$$G=Q\begin{bmatrix} E_r & G_{12} \\ G_{21} & G_{22} \end{bmatrix}P$$

因此

$$AGA=P^{-1}\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}Q^{-1}Q\begin{bmatrix} E_r & G_{12} \\ G_{21} & G_{22} \end{bmatrix}PP^{-1}\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}Q^{-1}=P^{-1}\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}Q^{-1}=A$$

由定理 8.2.2 可知，对任意 $A\in C^{m\times n}$，它的减号逆 A^- 总存在，并且

$$A^-=Q\begin{bmatrix} E_r & G_{12} \\ G_{21} & G_{22} \end{bmatrix}P$$

由 G_{12}，G_{21}，G_{22} 的任意性可知 A^- 不唯一. 定理 8.2.2 也告诉了我们求 A^- 的方法：

(1) 求使得 $PAQ=\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$ 的非奇异矩阵 P，Q；

(2) 写出 A 的减号逆 $A^-=Q\begin{bmatrix} I_r & G_{12} \\ G_{21} & G_{22} \end{bmatrix}P$.

由于

$$\begin{pmatrix} P & 0 \\ 0 & E \end{pmatrix}\begin{pmatrix} A & E \\ E & * \end{pmatrix}\begin{pmatrix} Q & 0 \\ 0 & E \end{pmatrix}=\begin{pmatrix} PA & P \\ E & * \end{pmatrix}\begin{pmatrix} Q & 0 \\ 0 & E \end{pmatrix}=\begin{pmatrix} PAQ & A \\ Q & * \end{pmatrix}$$

以 A 为基准，对矩阵 $\begin{pmatrix} A & E \\ E & * \end{pmatrix}$ 进行初等变换，E 的位置记录了对 A 进行变换的过程.

【例 8.2.1】 求 $A=\begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & 1 \\ 3 & 3 & 3 \end{bmatrix}$ 的减号逆 A^-.

解

$$\begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 0 \\ 2 & 2 & 1 & 0 & 1 & 0 \\ 3 & 3 & 3 & 0 & 0 & 1 \\ 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & * & \\ 0 & 0 & 1 & & & \end{bmatrix}\sim\begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & -3 & -2 & 1 & 0 \\ 0 & 0 & -3 & -3 & 0 & 1 \\ 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & * & \\ 0 & 0 & 1 & & & \end{bmatrix}\sim\begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{2}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \\ 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & * & \\ 0 & 0 & 1 & & & \end{bmatrix}$$

$$\sim\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{2}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \\ 1 & -1 & -2 & & & \\ 0 & 1 & 0 & & * & \\ 0 & 0 & 1 & & & \end{bmatrix}\sim\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & \frac{2}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \\ 1 & -2 & -1 & & & \\ 0 & 0 & 1 & & * & \\ 0 & 1 & 0 & & & \end{bmatrix}$$

因此

$$A^- = \begin{pmatrix} 1 & -2 & -1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & g_{13} \\ 0 & 1 & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ \frac{2}{3} & -\frac{1}{3} & 0 \\ -1 & -1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{1}{3} - g_{31} + g_{32} & \frac{2}{3} + \frac{1}{3}g_{32} - g_{13} + 2g_{23} + g_{33} & g_{13} - 2g_{23} - g_{33} \\ g_{31} + \frac{2}{3}g_{32} - g_{33} & -\frac{1}{3}g_{32} - g_{33} & g_{33} \\ \frac{2}{3} - g_{23} & -\frac{1}{3} - g_{23} & g_{23} \end{pmatrix}$$

其中 g_{31}，g_{32}，g_{13}，g_{23}，g_{33}是任意常数.

特别地，$A^- = \begin{pmatrix} -\frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 \\ \frac{2}{3} & -\frac{1}{3} & 0 \end{pmatrix}$就是 A 的一个减号逆.

定理 8.2.3 设 A^- 是 $A \in C^{m\times n}$ 的一个减号逆，则对任意

$$V \in C^{n\times m}, G = A^- + V - A^- AVAA^- \tag{8.2.2}$$

是 A 的减号逆，且 A 的任意一个减号逆都可以表示成式（8.2.2）的形式.

证 若 A^- 是 A 的一个减号逆，$G=A^-+V-A^-AVAA^-$，则

$$AGA = A(A^- + V - A^- AVAA^-)A = AA^- A + AVA - AA^- AVAA^- A$$
$$= A + AVA - AVA = A$$

故 $G=A^-+V-A^-AVAA^-$ 是 A 的减号逆.

另一方面，若 G 是 A 的任意一个减号逆，则 $A（G-A^-）A=A-A=0$，于是

$$G = A^- + (G - A^-) - A^- A(G - A^-)AA^-$$

取 $V=G-A^-$，则有 $G=A^-+V-A^-AVAA^-$.

定理 8.2.4 设 A^- 是 $A \in C^{m\times n}$ 的一个减号逆，则对任意 V，$W \in C^{n\times m}$，

$$G = A^- + V(E_m - AA^-) + (E_n - A^- A)W \tag{8.2.3}$$

是 A 的减号逆，且 A 的任意一个减号逆都可以表示成式（8.2.3）的形式.

证 若 A^- 是 A 的一个减号逆，$G=A^-+V(E_m-AA^-)+(E_n-A^-A)W$，易验证 $AGA=A$，从而 G 是 A 的一个减号逆.

另一方面，A^-，G 都是 A 的减号逆，则 $AA^-A=A$，$AGA=A$.

$$G = A^- + (G - A^-)$$
$$= A^- + (G - A^-) - (G - A^-)AA^- + GAA^- - A^- AA^-$$
$$= A^- + (G - A^-) - (G - A^-)AA^- + GAA^- - A^- AGAA^-$$
$$= A^- + (G - A^-)(E_m - AA^-) + (E_n - A^- A)GAA^-$$

取 $V=G-A^-$，$W=GAA^-$，则有 $G=A^-+V（E_m-AA^-）+（E_n-A^-A）W$.

定理 8.2.3，定理 8.2.4 给出了在已知一个减号逆的条件下，求其余减号逆的方法. 下面介绍广义逆矩阵 A^- 的性质.

定理 8.2.5 设 A 是 $m\times n$ 矩阵，k 是常数，则

(1) $\text{rank}(A^-)\geqslant\text{rank}(A)$；

(2) AA^-，A^-A 都是幂等阵，且

$$\text{rank}(AA^-)=\text{rank}(A^-A)=\text{rank}(A);$$

(3) A 非奇异时，$A\{1\}$ 只有唯一元素 A^{-1}；

(4) $(kA)^-=\begin{cases}\frac{1}{k}A^-, & k\neq 0\\ B, & k=0\end{cases}$，其中 $B\in\mathbf{C}^{n\times m}$ 是任意矩阵；

(5) $(A^-)^H=(A^H)^-$，$(A^-)^T=(A^T)^-$；

(6) $R(AA^-)=R(A),N(A^-A)=N(A),R((A^-A)^H)=R(A^H)$.

证 (1) 由 $AA^-A=A$ 有

$$\text{rank}(A)=\text{rank}(AA^-A)\leqslant\text{rank}(AA^-)\leqslant\text{rank}(A^-)$$

(2) 由 $AA^-A=A$ 有

$$\text{rank}(A)=\text{rank}(AA^-A)\leqslant\text{rank}(AA^-)\leqslant\text{rank}(A)$$

$$\text{rank}(A)=\text{rank}(AA^-A)\leqslant\text{rank}(A^-A)\leqslant\text{rank}(A)$$

因此

$$\text{rank}(AA^-)=\text{rank}(A^-A)=\text{rank}(A)$$

(3) $A^{-1}=A^{-1}AA^{-1}=A^{-1}AA^-AA^{-1}=A^-$.

(4) 利用矩阵的运算法则可直接验证其成立.

(5) 因为 $A^H(A^-)^HA^H=(AA^-A)^H=A^H$，由定义知，$(A^H)^-=(A^-)^H$，同理有 $(A^-)^T=(A^T)^-$.

(6) $R(A)\supset R(AA^-)\supset R(AA^-A)=R(A)$

$N(A)\subset N(A^-A)\subset N(AA^-A)=N(A)$

$R(A^H)\supset R(A^H(A^-)^H)=R((A^-A)^H)\supset R((AA^-A)^H)=R(A^H)$.

定理 8.2.6 (1) $AB(AB)^-A=A$ 的充分必要条件是 $\text{rank}(AB)=\text{rank}(A)$；

(2) $B(AB)^-AB=B$ 的充分必要条件是 $\text{rank}(AB)=\text{rank}(B)$.

证 (1) 充分性. 由 $\text{rank}(AB)=\text{rank}(A)$ 和 $R(AB)\subset R(A)$ 有，$R(AB)=R(A)$. 又因 $AB(AB)^-$ 是幂等阵，因此它是沿 M 向 L 的投影变换的矩阵，其中

$$L\oplus M=\mathbf{C}^n,L=R(AB(AB)^-)=R(AB)=R(A)$$

故

$$AB(AB)^-A=A$$

必要性. 因为 $AB(AB)^-A=A$，所以

$$\text{rank}(A)=\text{rank}(AB(AB)^-A)\leqslant\text{rank}(AB)\leqslant\text{rank}(A)$$

即 $\text{rank}(AB)=\text{rank}(A)$.

类似可证 (2).

定理 8.2.7 设 A，B 是 $m\times n$ 矩阵，非奇异矩阵 P，Q 使得 $B=PAQ$，则

$$B\{1\}=\{Q^{-1}A^-P^{-1}\mid A^-\in A\{1\}\}$$

证 设 A^- 是 A 的一个减号逆. 记 $G=Q^{-1}A^-P^{-1}$，则有

$$BGB=PAQQ^{-1}A^-P^{-1}PAQ=PAA^-AQ=PAQ=B$$

从而 $G\in B\{1\}$，即 $\{Q^{-1}A^{-}P^{-1}\mid A^{-}\in A\{1\}\}\subset B\{1\}$. 反之，若 $G\in B\{1\}$，则 $BGB=B$. 因此

$$PAQGPAQ = PAQ \quad , \quad AQGPA = A$$

由此有 $QGP\in A\{1\}$，$G=Q^{-1}QGPP^{-1}$，即 $B\{1\}\subset\{Q^{-1}A^{-}P^{-1}\mid A^{-}\in A\{1\}\}$.

综上所证，$B\{1\}=\{Q^{-1}A^{-}P^{-1}|A^{-}\in A\{1\}\}$.

定义 8.2.2 设 $A\in C^{m\times n}$.

(1) 若矩阵 $G\in C^{n\times m}$ 使得 $AG=E_m$，则称 G 是 A 的右逆；记作 A_R^-；

(2) 若矩阵 $G\in C^{n\times m}$ 使得 $GA=E_n$，则称 G 是 A 的左逆，记作 A_L^-.

定理 8.2.8 设 $A\in C^{m\times n}$.

(1) A 有右逆的充分必要条件是 A 行满秩，且 $A^H(AA^H)^{-1}$ 是 A_R^- 中的一个；

(2) A 有左逆的充分必要条件是 A 列满秩，且 $(A^HA)^{-1}A^H$ 是 A_L^- 中的一个.

证 (略).

定理 8.2.9 设 $A\in C^{m\times n}$.

(1) A^- 是 A 的右逆的充分必要条件是 rank $(A)=m$；

(2) A^- 是 A 的左逆的充分必要条件是 rank $(A)=n$.

证 (1) 必要性. 设 A^- 是 A 的右逆，则 $AA^-=E_m$，因此

$$m = \mathrm{rank}(AA^-) \leqslant \mathrm{rank}(A) \leqslant m$$

故 rank $(A)=m$.

充分性. 若 rank $(A)=m$，则 $n>m$，且存在非奇异矩阵 P，Q，使得 $PAQ=(E_m,0)$. 由定理 8.2.2 可知，A 的任意一个减号逆具有如下形式

$$A^- = Q\begin{pmatrix} E_m \\ G_{21} \end{pmatrix}P$$

其中 G_{21} 是 $(n-m)\times m$ 阶矩阵. 因此

$$AA^- = P^{-1}(E_m,0)Q^{-1}Q\begin{pmatrix} E_m \\ G_{21} \end{pmatrix}P = E_m$$

即 A^- 是 A 的右逆.

同理可证 (2).

推论 8.2.1 (1) 若存在非奇异阵 $Q=(Q_1,Q_2)$，使得 $AQ=(E_m,0)_{m\times n}$，则 $A^-=(Q_1,Q_2)\begin{pmatrix} E_m \\ G_{21} \end{pmatrix}=Q_1+Q_2G_{21}$ 是 A 的右逆.

(2) 若存在非奇异阵 $P=\begin{pmatrix} P_1 \\ P_2 \end{pmatrix}$，使得 $PA=\begin{pmatrix} E_n \\ 0 \end{pmatrix}_{m\times n}$，则 $A^-=(E_m,G_{12})\begin{pmatrix} P_1 \\ P_2 \end{pmatrix}=P_1+G_{12}P_2$ 是 A 的左逆.

下面我们来讨论方程组的求解问题.

定理 8.2.10 线性齐次方程组 $Ax=0$ 的通解是

$$x = (E_n - A^-A)y \quad , \quad y \in \mathbf{R}^n$$

证 实际上，只需证明 $N(A)=R(E_n-A^-A)$.

对任意 $x\in R(E_n-A^-A)$，存在 $y\in\mathbf{R}^n$，使得 $x=(E_n-A^-A)y$，因此

$$Ax = A(E_n - A^-A)y = Ay - AA^-Ay = Ay - Ay = 0$$

从而 $\mathbf{R}(E_n - A^- A) \subset \mathbf{N}(A)$.

所以 $\dim(\mathbf{N}(A)) = n - \operatorname{rank}(A)$. 由定理 8.2.5 有，$A^- A$ 是幂等阵，且

$$\operatorname{rank}(AA^-) = \operatorname{rank}(A^- A) = \operatorname{rank}(A)$$

因此

$$\begin{aligned}\dim(\mathbf{R}(E_n - A^- A)) &= \operatorname{rank}(E_n - A^- A) \\ &= n - \dim(\mathbf{R}(A^- A)) \\ &= n - \dim(\mathbf{R}(A)) \\ &= n - \operatorname{rank}(A)\end{aligned}$$

因此 $\dim(\mathbf{N}(A)) = \dim(\mathbf{R}(E_n - A^- A))$. 故 $\mathbf{N}(A) = \mathbf{R}(E_n - A^- A)$.

定理 8.2.11 若 A^- 是 A 的一个减号逆，则相容方程组 $Ax=b$ 的通解是

$$x = A^- b + (E_n - A^- A)y \quad , \quad y \in \mathbf{R}^n$$

证 由定理 8.2.12，只需证明 $x = A^- b$ 是 $Ax=b$ 的特解.

由于 $Ax=b$ 相容方程组，因此存在 $y \in \mathbf{R}^n$，使得 $b = Ay$. 故

$$Ax = AA^- b = AA^- Ay = Ay = b$$

即 $x = A^- b$ 是 $Ax=b$ 的特解.

【例 8.2.2】 求方程 $\begin{cases} x_1 + x_2 + 2x_3 = 4 \\ 2x_1 + 2x_2 + x_3 = 5 \\ 3x_1 + 3x_2 + 3x_3 = 9 \end{cases}$ 的通解.

解 方程组的系数矩阵 $A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 1 \\ 3 & 3 & 3 \end{pmatrix}$. 由例 8.2.1 知，$A^- = \begin{pmatrix} -\frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 \\ \frac{2}{3} & -\frac{1}{3} & 0 \end{pmatrix}$. 因此，

该方程组的通解为

$$x = \begin{pmatrix} -\frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 \\ \frac{2}{3} & -\frac{1}{3} & 0 \end{pmatrix} \begin{pmatrix} 4 \\ 5 \\ 9 \end{pmatrix} + \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} -\frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 \\ \frac{2}{3} & -\frac{1}{3} & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 1 \\ 3 & 3 & 3 \end{pmatrix} \right) \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$= \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} -y_2 \\ y_2 \\ 0 \end{pmatrix}$$

其中 $y_2 \in \mathbf{R}$.

§8.3 自反广义逆

对于非奇异矩阵 A，其逆矩阵 A^{-1} 满足 $(A^{-1})^{-1} = A$. 对一般的广义逆矩阵 A^-，$(A^-)^- \neq A$. 如 $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$，$G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in A\{1\}$，$G^- = G \neq A$. 但对某些特殊的广义逆矩阵，这种自反性质却成立.

定义 8.3.1 设$A\in C^{m\times n}$，若矩阵$G\in C^{n\times m}$使得$AGA=A$，$GAG=G$，则称G是A的自反广义逆，记作A_r^-，或记$G\in A\{1,2\}$.

定理 8.3.1 设$A\in C^{m\times n}$，则A^-是A的自反广义逆的充分必要条件是 rank(A^-) = rank(A).

证 必要性. 设A^-是A的自反广义逆. 由$AA^-A=A$有

$$\text{rank}(A)=\text{rank}(AA^-A)\leqslant\text{rank}(AA^-)\leqslant\text{rank}(A^-)$$

由$A^-AA^-=A^-$有

$$\text{rank}(A^-)=\text{rank}(A^-AA^-)\leqslant\text{rank}(AA^-)\leqslant\text{rank}(A)$$

因此 rank(A^-) =rank(A).

充分性. 由$AA^-A=A$有

$$AA^- - AA^-AA^- = AA^-(E-AA^-)=0$$

因此$E-AA^-$的列向量是$AA^-x=0$的解.

又由于 rank(AA^-) =rank(A) =rank(A^-)，因此$AA^-x=0$与$A^-x=0$是同解方程组，故

$$A^-(E-AA^-)=0$$

即$A^-AA^-=A^-$，从而A^-是A的自反广义逆.

由定理 8.2.2 知，如果非奇异矩阵$P\in C^{m\times m}$，$Q\in C^{n\times n}$使得

$$PAQ=\begin{pmatrix}E_r & 0\\ 0 & 0\end{pmatrix}$$

则

$$A^-=Q\begin{pmatrix}E_r & G_{12}\\ G_{21} & G_{22}\end{pmatrix}P$$

其中G_{12}，G_{21}，G_{22}分别是$r\times(m-r)$，$(n-r)\times r$，$(n-r)\times(m-r)$的任意矩阵.

rank(A^-) =rank(A) 的充分必要条件是$G_{22}=G_{21}G_{12}$. 因此有

定理 8.3.2 设A是$m\times n$矩阵，rank$(A)=r$，非奇异矩阵$P\in C^{m\times m}$，$Q\in C^{n\times n}$使得

$$PAQ=\begin{pmatrix}E_r & 0\\ 0 & 0\end{pmatrix}$$

则$G\in C^{n\times m}$是A的自反广义逆的充分必要条件是

$$G=Q\begin{pmatrix}E_r & G_{12}\\ G_{21} & G_{21}G_{12}\end{pmatrix}P$$

其中G_{12}，G_{21}分别是$r\times(m-r)$，$(n-r)\times r$的任意矩阵.

定理 8.3.3 设$A\in C^{m\times n}$，rank$(A)=r>0$，$A=BC$是A的满秩分解，则$G=C_R^-B_L^-$是A的自反广义逆，且A的任何自反广义逆都可以表示成该形式.

证 由于 rank$(A)=r>0$，$A=BC$是A的满秩分解，因此B行满秩，C列满秩，从而C的右逆C_R^-存在，B的左逆B_L^-存在.

$$AGA=BCC_R^-B_L^-BC=BC=A,$$
$$GAG=C_R^-B_L^-BCC_R^-B_L^-=C_R^-B_L^-=G,$$

因此$G=C_R^-B_L^-$是A的自反广义逆.

设A_r^-是A的任意一个自反广义逆，则

$$BCA_r^- BC = AA_r^- A = A = BC \quad , \quad CA_r^- B = E_r$$

由定义知，CA_r^- 是 B 的左逆，$A_r^- B$ 是 C 的右逆. 又

$$A_r^- = A_r^- AA_r^- = A_r^- BCA_r^- = (A_r^- B)(CA_r^-)$$

因此可得，A 的任何自反广义逆都可以表示成 C 的右逆右乘 B 的左逆的形式.

§8.4 广义逆 A_m^-

在前面讲过，对于相容方程 $Ax=b$，若能找到 A 的一个减号逆 A^-，则它的通解可表示成

$$x = A^- b + (E_n - A^- A)y \quad , \quad y \in C^n$$

在这些解中 2-范数最小的解 x_0，即满足

$$\|x_0\|_2 = \min\{\|x\|_2 \mid Ax = b\}$$

的解 x_0，我们称为相容方程 $Ax=b$ 的极小范数解. 现在，我们考虑的问题是，是否存在矩阵 G，使得对任意的 $b\in R(A)$，$x=Gb$ 都是方程 $Ax=b$ 的极小范数解？如果 G 存在，它应有什么特征？

引理 8.4.1 对任意的 u，$y\in C^n$，$2\mathrm{Re}[u^H(GA)^H(E_n-GA)y]+\|(E_n-GA)y\|_2^2\geqslant 0$ 都成立的充分必要条件是

$$(GA)^H(E_n - GA) = 0$$

证 若 $(GA)^H(E_n-GA)=0$，显然，任意的 u，$y\in C^n$，

$$2\mathrm{Re}[u^H(GA)^H(E_n - GA)y] + \|(E_n - GA)y\|_2^2 \geqslant 0$$

反之，设对任意的 u，$y\in C^n$，都有 $2\mathrm{Re}[u^H(GA)^H(E_n-GA)y]+\|(E_n-GA)y\|_2^2\geqslant 0$. 如果 $(GA)^H(E_n-GA)\neq 0$，则它至少有一个元素不为零，不妨设其 $(i，j)$ 处元素 $b_{ij}\neq 0$. 取 $y=e_j=(0，0，\cdots，0，1，0，\cdots，0)^T$，则

$$(GA)^H(E_n - GA)e_j = b_{ij}e_i \neq 0$$

因此

$$\|(E_n - GA)e_j\|_2^2 > 0$$

取 $\tilde{u}=-\dfrac{\|(E_n-GA)\ y\|_2^2}{\bar{b}_{ij}}e_i$，则

$$2\mathrm{Re}[\tilde{u}^H(GA)^H(E_n - GA)e_j] + \|(E_n - GA)e_j\|_2^2 = -\|(E_n - GA)e_j\|_2^2 < 0$$

这与条件矛盾. 因此 $(GA)^H(E_n-GA)=0$.

定理 8.4.1 设 $A\in C^{m\times n}$，$G\in C^{n\times m}$. 对任意的 $b\in R(A)$，$x=Gb$ 都是方程 $Ax=b$ 的极小范数解的充分必要条件是 G 满足

$$AGA = A \quad , \quad (GA)^H = GA$$

证 对任意的 $b\in R(A)$，$x=Gb$ 都是方程 $Ax=b$ 的解的充分必要条件是 G 满足 $AGA=A$. 因此要证明定理成立，只需证明若 $G\in A\{1\}$，对任意的 $b\in R(A)$，$x=Gb$ 都是方程 $Ax=b$ 的极小范数解的充分必要条件是 G 满足 $(GA)^H=GA$.

由于相容方程 $Ax=b$ 的通解为

$$x = Gb + (E_n - GA)y \quad , \quad y \in C^n$$

因此任意的 $b\in R(A)$，$x=Gb$ 都是方程 $Ax=b$ 的极小范数解充分必要条件是

$$\|Gb\|_2 \leqslant \|Gb+(E_n-GA)y\|_2 \quad 对任意的\ b\in R(A), y\in C^n\ 成立$$

充分必要条件是

$$\|GAu\|_2 \leqslant \|GAu+(E_n-GA)y\|_2 \quad 对任意的\ u, y\in C^n\ 成立$$

充分必要条件是

$$2\mathrm{Re}[u^H(GA)^H(E_n-GA)y]+\|(E_n-GA)y\|_2^2 \geqslant 0 \quad 对任意的\ u, y\in C^n\ 成立$$

充分必要条件是

$$(GA)^H(E_n-GA)=0$$

即

$$(GA)^H=(GA)^HGA;$$

充分必要条件是 $(GA)^H=GA$ (注意到 $AGA=A$).

定义 8.4.1 对任意的 $b\in R(A)$，矩阵 G 使得 $x=Gb$ 都是方程 $Ax=b$ 的极小范数解，则称 G 为 A 的极小范数广义逆，记为 A_m^-，$G\in A\{1,4\}$.

定理 8.4.2 $G=A^H(AA^H)^-\in A\{1,2,4\}$，从而它是 A 的极小范数广义逆.

证 $(GA)^H=(A^H(AA^H)^-A)^H=A^H[(AA^H)^-]^HA=A^H[(AA^H)^H]^-A=A^H(AA^H)^-A=GA$

$$\begin{aligned}
&(AGA-A)(AGA-A)^H\\
&=[AA^H(AA^H)^-A-A][AA^H(AA^H)^-A-A]^H\\
&=[AA^H(AA^H)^-A-A][A^H(AA^H)^-AA^H-A^H]\\
&=[AA^H(AA^H)^-AA^H-AA^H][(AA^H)^-AA^H-E]\\
&=[AA^H-AA^H][(AA^H)^-AA^H-E]\\
&=0
\end{aligned}$$

因此 $AGA=A$，即 $G=A^H(AA^H)^-\in A\{1\}$，故 $\mathrm{rank}(A)\leqslant\mathrm{rank}(G)$.

另一方面

$$\mathrm{rank}(G)=\mathrm{rank}(A^H(AA^H)^-)\leqslant\mathrm{rank}(A^H)=\mathrm{rank}(A)$$

因此 $\mathrm{rank}(G)=\mathrm{rank}(A)$. 由定理 8.3.1 知，$G\in A\{2\}$.

综上所证，$G=A^H(AA^H)^-\in A\{1,2,4\}$.

【例 8.4.1】 设 $A=\begin{pmatrix}1&2&-1&1\\0&-1&2&1\\1&1&1&2\end{pmatrix}$，求 A_m^-.

解 $AA^H=\begin{pmatrix}7&-3&4\\-3&6&3\\4&3&7\end{pmatrix}$

$$\begin{pmatrix}7&-3&4&1&0&0\\-3&6&3&0&1&0\\4&3&7&0&0&1\\1&0&0&&&\\0&1&0&&*&\\0&0&1&&&\end{pmatrix}\sim\begin{pmatrix}1&9&10&1&2&0\\-1&2&1&0&1/3&0\\0&11&11&0&4/3&1\\1&0&0&&&\\0&1&0&&*&\\0&0&1&&&\end{pmatrix}\sim\begin{pmatrix}1&9&10&1&2&0\\0&11&11&1&7/3&0\\0&0&0&-1&-1&1\\1&0&0&&&\\0&1&0&&*&\\0&0&1&&&\end{pmatrix}$$

$$\sim\begin{pmatrix}1&0&0&1&2&0\\0&1&1&1/11&7/33&0\\0&0&0&-1&-1&1\\1&-9&-10&&&\\0&1&0&&*&\\0&0&1&&&\end{pmatrix}\sim\begin{pmatrix}1&0&0&1&2&0\\0&1&0&1/11&7/33&0\\0&0&0&-1&-1&1\\1&-9&-1&&&\\0&1&-1&&*&\\0&0&1&&&\end{pmatrix}$$

因此

$$\begin{pmatrix}1&-9&-1\\0&1&-1\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}\begin{pmatrix}1&2&0\\1/11&7/33&0\\-1&-1&1\end{pmatrix}=\begin{pmatrix}2/11&1/11&0\\1/11&7/33&0\\0&0&0\end{pmatrix}$$

是 $\boldsymbol{AA}^{\mathrm{H}}$ 的一个（自反）广义逆.

$$\boldsymbol{G}=\begin{pmatrix}1&2&-1&1\\0&-1&2&1\\1&1&1&2\end{pmatrix}^{\mathrm{H}}\begin{pmatrix}2/11&1/11&0\\1/11&7/33&0\\0&0&0\end{pmatrix}=\begin{pmatrix}2/11&1/11&0\\3/11&-1/33&0\\0&1/3&0\\3/11&10/33&0\end{pmatrix}$$ 是 $\boldsymbol{A}$ 的一个极小范数广义逆.

定理 8.4.3 设 $\boldsymbol{A}_m^- \in \boldsymbol{A}\{1,4\}$，则 $\boldsymbol{G}\in\boldsymbol{A}\{1,4\}$ 的充分必要条件是 $\boldsymbol{GA}=\boldsymbol{A}_m^-\boldsymbol{A}$.

证 若 $\boldsymbol{GA}=\boldsymbol{A}_m^-\boldsymbol{A}$，由于 $\boldsymbol{A}_m^-\in\boldsymbol{A}\{1,4\}$，

$$\boldsymbol{AGA}=\boldsymbol{AA}_m^-\boldsymbol{A}=\boldsymbol{A},(\boldsymbol{GA})^{\mathrm{H}}=(\boldsymbol{A}_m^-\boldsymbol{A})^{\mathrm{H}}=\boldsymbol{A}_m^-\boldsymbol{A}=\boldsymbol{GA}$$

因此 $\boldsymbol{G}\in\boldsymbol{A}\{1,4\}$.

反之，若 $\boldsymbol{G}\in\boldsymbol{A}\{1,4\}$，则有

$$\boldsymbol{GA}=(\boldsymbol{GA})^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}_m^-)^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}=(\boldsymbol{A}_m^-\boldsymbol{A})^{\mathrm{H}}(\boldsymbol{GA})^{\mathrm{H}}=\boldsymbol{A}_m^-\boldsymbol{AGA}=\boldsymbol{A}_m^-\boldsymbol{A}$$

定理 8.4.4 设 $\boldsymbol{A}_m^-\in\boldsymbol{A}\{1,4\}$，则 $\boldsymbol{G}\in\boldsymbol{A}\{1,4\}$ 的充分必要条件是 $\boldsymbol{G}$ 可以表示成

$$\boldsymbol{G}=\boldsymbol{A}_m^-+\boldsymbol{Z}(\boldsymbol{E}_n-\boldsymbol{AA}_m^-)$$

其中 $\boldsymbol{Z}\in\mathbf{C}^{n\times m}$.

证 设 $\boldsymbol{A}_m^-\in\boldsymbol{A}\{1,4\}$. 若 $\boldsymbol{G}=\boldsymbol{A}_m^-+\boldsymbol{Z}(\boldsymbol{E}_n-\boldsymbol{AA}_m^-)$，则有

$$\boldsymbol{GA}=\boldsymbol{A}_m^-\boldsymbol{A}+\boldsymbol{Z}(\boldsymbol{E}_n-\boldsymbol{AA}_m^-)\boldsymbol{A}=\boldsymbol{A}_m^-\boldsymbol{A}+\boldsymbol{Z}(\boldsymbol{A}-\boldsymbol{AA}_m^-\boldsymbol{A})=\boldsymbol{A}_m^-\boldsymbol{A}$$

由定理 8.4.3，$\boldsymbol{G}\in\boldsymbol{A}\{1,4\}$.

反之，若 $\boldsymbol{G}\in\boldsymbol{A}\{1,4\}$，由定理 8.4.3 有，$\boldsymbol{GA}=\boldsymbol{A}_m^-\boldsymbol{A}$，即 $(\boldsymbol{G}-\boldsymbol{A}_m^-)\boldsymbol{A}=0$. 因而

$$\begin{aligned}\boldsymbol{G}&=\boldsymbol{A}_m^-+(\boldsymbol{G}-\boldsymbol{A}_m^-)\\&=\boldsymbol{A}_m^-+(\boldsymbol{G}-\boldsymbol{A}_m^-)-(\boldsymbol{G}-\boldsymbol{A}_m^-)\boldsymbol{A}\\&=\boldsymbol{A}_m^-+(\boldsymbol{G}-\boldsymbol{A}_m^-)-(\boldsymbol{G}-\boldsymbol{A}_m^-)\boldsymbol{AA}_m^-\\&=\boldsymbol{A}_m^-+(\boldsymbol{G}-\boldsymbol{A}_m^-)(\boldsymbol{E}_{11}-\boldsymbol{AA}_m^-)\end{aligned}$$

取 $\boldsymbol{Z}=\boldsymbol{G}-\boldsymbol{A}_m^-$，则 $\boldsymbol{G}=\boldsymbol{A}_m^-+\boldsymbol{Z}(\boldsymbol{E}_n-\boldsymbol{AA}_m^-)$.

定理 8.4.5 相容方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 的极小范数解唯一存在.

证 由于 $\boldsymbol{A}$ 的极小范数广义逆 $\boldsymbol{G}$ 存在，因此 $\boldsymbol{x}_0=\boldsymbol{Gb}$ 是 $\boldsymbol{Ax}=\boldsymbol{b}$ 的极小范数解.

若 $\boldsymbol{y}$ 是 $\boldsymbol{Ax}=\boldsymbol{b}$ 的极小范数解，则存在 $\boldsymbol{z}\in\mathbf{C}^n$，使得 $\boldsymbol{y}=\boldsymbol{Gb}+(\boldsymbol{E}_n-\boldsymbol{GA})\boldsymbol{z}$，且

$$\begin{aligned}\|\boldsymbol{Gb}\|^2=\|\boldsymbol{y}\|^2&=\|\boldsymbol{Gb}+(\boldsymbol{E}_n-\boldsymbol{GA})\boldsymbol{z}\|^2,\\&=\|\boldsymbol{Gb}\|_2^2+2\mathrm{Re}[\boldsymbol{u}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}(\boldsymbol{E}_n-\boldsymbol{GA})\boldsymbol{z}]+\|(\boldsymbol{E}_n-\boldsymbol{GA})\boldsymbol{z}\|_2^2\end{aligned}$$

$$= \|Gb\|_2^2 + \|(E_n - GA)z\|_2^2$$

因此

$$\|(E_n - GA)z\|_2 = 0$$

即

$$(E_n - GA)z = 0$$

因此 $y=Gb$.

【例 8.4.2】 设 $A=\begin{pmatrix}1 & 2 & -1 & 1\\ 0 & -1 & 2 & 1\\ 1 & 1 & 1 & 2\end{pmatrix}$，$b=\begin{pmatrix}3\\2\\5\end{pmatrix}$，求 $Ax=b$ 的极小范数解.

解 由例 8.4.1 有 $G=\frac{1}{33}\begin{pmatrix}6 & 3 & 0\\ 9 & -1 & 0\\ 0 & 11 & 0\\ 9 & 10 & 0\end{pmatrix}$是 A 的一个极小范数广义逆. 因此 $Ax=b$ 的极小范数解是

$$x = Gb = \frac{1}{33}\begin{pmatrix}6 & 3 & 0\\ 9 & -1 & 0\\ 0 & 11 & 0\\ 9 & 10 & 0\end{pmatrix}\begin{pmatrix}3\\2\\5\end{pmatrix} = \frac{1}{33}\begin{pmatrix}24\\25\\22\\47\end{pmatrix}$$

§8.5 广义逆 A_l^-

设 $Ax=b$ 是不相容方程，它在通常意义下无解. 现要求这样的解 x_0，使得

$$\|Ax_0 - b\|_2 = \min\{\|Ax - b\|_2 \mid x \in \mathbf{C}^n\}$$

这样的解 x_0，称为矛盾方程 $Ax=b$ 的最小二乘解. 现在，我们考虑的问题是，是否存在矩阵 G，使得对任意的 $b\in\mathbf{C}^n$，$x=Gb$ 都是方程 $Ax=b$ 的最小二乘解？如果 G 存在，它应有什么特征？

定理 8.5.1 设 $A\in\mathbf{C}^{m\times n}$，$G\in\mathbf{C}^{n\times m}$. 对任意的 $b\in\mathbf{C}^n$，$x=Gb$ 都是方程 $Ax=b$ 的最小二乘解的充分必要条件是 G 满足

$$AGA = A \quad , \quad (AG)^{\mathrm{H}} = AG$$

证

$$\begin{aligned}\|Ax - b\|_2^2 &= \|AGb - b + Ax - AGb\|_2^2\\ &= \|AGb - b\|_2^2 + \|Ax - AGb\|_2^2 + 2\mathrm{Re}[(AGb - b)^{\mathrm{H}}(Ax - AGb)]\end{aligned}$$

因此

$$\begin{aligned}\|Ax - b\|_2^2 - \|AGb - b\|_2^2 &= \|Ax - AGb\|_2^2 + 2\mathrm{Re}[(AGb - b)^{\mathrm{H}}(Ax - AGb)]\\ &= \|A(x - Gb)\|_2^2 + 2\mathrm{Re}[b^{\mathrm{H}}(AG - E)^{\mathrm{H}}A(x - Gb)]\end{aligned}$$

因此对任意的 $b\in\mathbf{C}^m$，$y\in\mathbf{C}^n$，$x_0=Gb$ 都是方程 $Ax=b$ 的最小二乘解的充分必要条件是

$$\|Ax - b\|_2^2 - \|AGb - b\|_2^2 \geqslant 0, \text{对任意的 } b \in \mathbf{C}^m, x \in \mathbf{C}^n \text{ 成立}$$

充分必要条件是

$\|A(x-Gb)\|_2^2+2\mathrm{Re}[b^H(AG-E)^H A(x-Gb)]\geqslant 0$，对任意的 $b\in C^m$，$x\in C^n$ 成立充分必要条件是

$$(E_m-AG)^H A=0$$

即

$$(AG)^H A=A$$

若 $(AG)^H A=A$，则有 $(AG)^H AG=AG$，$(AG)^H AG=(AG)^H$，因此

$$(AG)^H=AG,AGA=(AG)^H A=A$$

反之，若 $(AG)^H=AG$，$AGA=A$，则 $(AG)^H A=AGA=A$.

充分必要条件是 $(GA)^H=GA$（注意到 $AGA=A$）.

这就证明了对任意的 $b\in C^n$，$x=Gb$ 都是方程 $Ax=b$ 的最小二乘解的充分必要条件是 G 满足 $AGA=A$，且 $(AG)^H=AG$.

定义 8.5.1 对任意的 $b\in C^m$，矩阵 G 使得 $x=Gb$ 都是方程 $Ax=b$ 的最小二乘解，则称 G 为 A 的最小二乘广义逆，记为 A_l^-，$G\in A\{1,3\}$.

定理 8.5.2 $G=(A^HA)^-A^H\in A\{1,2,3\}$，因而它是 A 的最小二乘广义逆.

证 $(AG)^H=[A(A^HA)^-A^H]^H=A[(A^HA)^-]^HA^H=A(A^HA)^-A^H=AG$

$$\begin{aligned}&(AGA-A)^H(AGA-A)\\&=[A(A^HA)^-A^HA-A]^H[A(A^HA)^-A^HA-A]\\&=[A^HA(A^HA)^-A^H-A^H][A(A^HA)^-A^HA-A]\\&=[A^HA(A^HA)^--E_m][A^HA(A^HA)^-A^HA-A^HA]\\&=[A^HA(A^HA)^--E_m][A^HA-A^HA]\\&=0\end{aligned}$$

因此 $AGA=A$. 故 $G=(A^HA)^-A^H\in A\{1,3\}$.

由减号逆的性质有 rank $(A)\leqslant$rank (G). 由 G 的构成有，rank $(G)\leqslant$rank (A). 故 rank $(G)=$rank (A). 由定理 8.3.1 知，$G\in A\{2\}$.

综上所证，$G=(A^HA)^-A^H\in A\{1,2,3\}$.

【例 8.5.1】 设 $A=\begin{pmatrix}1&0&1\\0&1&1\\1&-1&0\\1&1&2\end{pmatrix}$，求 A_l^-.

解 $A^HA=\begin{pmatrix}3&0&3\\0&3&3\\3&3&6\end{pmatrix}$

$$\begin{pmatrix}3&0&3&1&0&0\\0&3&3&0&1&0\\3&3&6&0&0&1\\1&0&0&&&\\0&1&0&&*&\\0&0&1&&&\end{pmatrix}\sim\begin{pmatrix}3&0&3&1&0&0\\0&3&3&0&1&0\\0&3&3&-1&0&1\\1&0&0&&&\\0&1&0&&*&\\0&0&1&&&\end{pmatrix}\sim\begin{pmatrix}3&0&3&1&0&0\\0&3&3&0&1&0\\0&0&0&-1&-1&1\\1&0&0&&&\\0&1&0&&*&\\0&0&1&&&\end{pmatrix}$$

$$\sim\begin{pmatrix}3&0&0&1&0&0\\0&3&0&0&1&0\\0&0&0&-1&-1&1\\1&0&-1&&&\\0&1&-1&&*&\\0&0&1&&&\end{pmatrix}\sim\begin{pmatrix}1&0&0&1/3&0&0\\0&1&0&0&1/3&0\\0&0&0&-1&-1&1\\1&0&-1&&&\\0&1&-1&&*&\\0&0&1&&&\end{pmatrix}$$

因此

$$\begin{pmatrix}1&0&-1\\0&1&-1\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}\begin{pmatrix}1/3&0&0\\0&1/3&0\\-1&-1&1\end{pmatrix}=\frac{1}{3}\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}$$

是 A^HA 的一个减号逆.

$$G=\frac{1}{3}\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}\begin{pmatrix}1&0&1\\0&1&1\\1&-1&0\\1&1&2\end{pmatrix}^H=\frac{1}{3}\begin{pmatrix}1&0&1&1\\0&1&-1&1\\0&0&0&0\end{pmatrix}$$

是 A 的一个最小二乘广义逆.

定理 8.5.3 设 $A_l^-\in A\{1,3\}$，则 $G\in A\{1,3\}$的充分必要条件是 $AG=AA_l^-$.

证 若 $AG=AA_l^-$，由于 $A_l^-\in A\{1,3\}$，有

$$AGA=AA_l^-A=A,(AG)^H=(AA_l^-)^H=AA_l^-=AG$$

因此 $G\in A\{1,3\}$.

反之，若 $G\in A\{1,3\}$，则有

$$AG=(AG)^H=G^HA^H=G^HA^H(A_l^-)^HA^H=(AG)^H(AA_l^-)^H=AGAA_l^-=AA_l^-$$

定理 8.5.4 设 $A_l^-\in A\{1,3\}$，则 $G\in A\{1,3\}$ 的充分必要条件是 G 可以表示成

$$G=A_l^-+(E_n-A_l^-A)Z$$

其中 $Z\in C^{n\times m}$.

证 设 $A_l^-\in A\{1,3\}$. 若 $G=A_l^-+(E_n-A_l^-A)Z$，则有

$$AG=AA_l^-+A(E_n-A_l^-A)Z=AA_l^-+(A-AA_l^-A)Z=AA_l^-$$

由定理 8.5.3，$G\in A\{1,3\}$.

反之，若 $G\in A\{1,3\}$，由定理 8.5.3 有，$AG=AA_l^-$，即 $A(G-A_l^-)=0$. 因而

$$\begin{aligned}G&=A_l^-+(G-A_l^-)\\&=A_l^-+(G-A_l^-)-A(G-A_l^-)\\&=A_l^-+(G-A_l^-)-A_l^-A(G-A_l^-)\\&=A_l^-+(E-AA_l^-)(G-A_l^-)\end{aligned}$$

取 $Z=G-A_l^-$，则 $G=A_l^-+(E_n-A_l^-A)Z$.

定理 8.5.5 矛盾方程 $Ax=b$ 的最小二乘解是

$$x=A_l^-b+(E_n-A_l^-A)y\quad,\quad y\in C^n$$

证 由于 $Ax=b$ 是矛盾方程，因此 $b\neq 0$. 由定理 8.5.4 知，矛盾方程 $Ax=b$ 的最小二乘解是

$$x=A_l^-b+(E_n-A_l^-A)Zb$$

由 Z 的任意性知，矛盾方程 $Ax=b$ 的最小二乘解是

$$x = A_l^- b + (E_n - A_l^- A)y \quad , \quad y \in C^n$$

【例 8.5.2】 设 $A=\begin{pmatrix}1&0&1\\0&1&1\\1&-1&0\\1&1&2\end{pmatrix}$，$b=\begin{pmatrix}1\\2\\3\\4\end{pmatrix}$，求矛盾方程 $Ax=b$ 的最小二乘解.

解 因为 $A_l^-=\frac{1}{3}\begin{pmatrix}1&0&1&1\\0&1&-1&1\\0&0&0&0\end{pmatrix}$是 A 的一个最小二乘广义逆. 矛盾方程 $Ax=b$ 的最小二乘解是

$$\begin{aligned}x&= A_l^- b + (E_n - A_l^- A)y\\
&=\frac{1}{3}\begin{pmatrix}1&0&1&1\\0&1&-1&1\\0&0&0&0\end{pmatrix}\begin{pmatrix}1\\2\\3\\4\end{pmatrix}+\left(\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}-\frac{1}{3}\begin{pmatrix}1&0&1&1\\0&1&-1&1\\0&0&0&0\end{pmatrix}\begin{pmatrix}1&0&1\\0&1&1\\1&-1&0\\1&1&2\end{pmatrix}\right)y\\
&=\frac{1}{3}\begin{pmatrix}1&0&1&1\\0&1&-1&1\\0&0&0&0\end{pmatrix}\begin{pmatrix}1\\2\\3\\4\end{pmatrix}+\begin{pmatrix}0&0&-1\\0&0&-1\\0&0&1\end{pmatrix}\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}\\
&=\frac{1}{3}\begin{pmatrix}8\\3\\0\end{pmatrix}+k\begin{pmatrix}1\\1\\-1\end{pmatrix}\end{aligned}$$

其中 k 是任意常数.

定理 8.5.6 若 A 列满秩，则 A 的最小二乘广义逆 G 唯一存在，且 $G=(A^HA)^{-1}A^H$.

证 先证 $G=(A^HA)^{-1}A^H$ 是 A 的最小二乘广义逆. 若 A 列满秩，则 $(A^HA)^{-1}$存在，

$$GA = (A^HA)^{-1}A^HA = E_n$$

因此 $(GA)^H=E_n=GA$，$AGA=AE_n=A$，故 $G\in A\{1,3\}$.

再证唯一性. 对 A 的任一最小二乘广义逆 A_l^-，由定理 8.5.3 有 $AG=AA_l^-$，因此

$$G-A_l^- = GA(G-A_l^-) = GAG - GAA_l^- = GAG - GAG = 0$$

故 $A_l^-=G$，即 A 的最小二乘广义逆 G 唯一.

推论 8.5.1 若 A 列满秩，则 $Ax=b$ 的最小二乘解 x 唯一存在，且 $x=(A^HA)^{-1}A^Hb$.

在曲线拟合和多元线性回归分析等问题中，常常要计算矛盾方程 $Ax=b$ 的最小二乘解，广义逆矩阵的理论使求矛盾方程 $Ax=b$ 的最小二乘解的方法标准化、规范化，整个求解过程归结为求 A 的最小二乘广义逆 A_l^-，而不用求误差平方和的极值等一套繁琐的过程.

曲线拟合的最小二乘问题可以描述为：根据已知的数据组 (x_i, y_i)，$(i=1,2,\cdots,n)$，选取一个近似函数 $y=f(x)$，使得

$$\sum_{i=1}^{n}(y_i-f(x_i))^2$$

最小. 这种求近似函数的方法称为曲线拟合的最小二乘法，函数 $y=f(x)$ 称为这组数据的

最小二乘函数. 常用的拟合函数是多项式，即

$$f(x) = \sum_{k=0}^{m} a_k x^k$$

【例 8.5.3】 有函数 $y=f(x)$，试用一个二次最小平方逼近多项式拟合它.

0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3.1950	3.2299	3.2532	3.2611	3.2516	3.2282	3.1807	3.1266	3.0594	3.9759

解 设二次最小平方逼近多项式为

$$P_2(x) = a_0 + a_1 x + a_2 x^2$$

则 a_0，a_1，a_2 是方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二乘解. 其中

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0.1 & 0.01 \\ 1 & 0.2 & 0.04 \\ 1 & 0.3 & 0.09 \\ 1 & 0.4 & 0.16 \\ 1 & 0.5 & 0.25 \\ 1 & 0.6 & 0.36 \\ 1 & 0.7 & 0.49 \\ 1 & 0.8 & 0.64 \\ 1 & 0.9 & 0.81 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 3.1950 \\ 3.2299 \\ 3.2532 \\ 3.2611 \\ 3.2516 \\ 3.2282 \\ 3.1807 \\ 3.1266 \\ 3.0594 \\ 3.9759 \end{pmatrix}$$

解得

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \boldsymbol{A}_l^- \boldsymbol{b} = \begin{pmatrix} 3.19510 \\ 0.44255 \\ -0.76531 \end{pmatrix}$$

§8.6 广义逆 A^+

由定理 8.5.5 知，方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二乘解是

$$\boldsymbol{x} = \boldsymbol{A}_l^- \boldsymbol{b} + (\boldsymbol{E}_n - \boldsymbol{A}_l^- \boldsymbol{A})\boldsymbol{y} \quad , \quad \boldsymbol{y} \in \mathbf{C}^n$$

方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二乘解一般不唯一. 方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 的所有最小二乘解中范数最小的解，称为方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 的极小最小二乘解. 现在的问题是：对任意 $\boldsymbol{b}\in\mathbf{C}^m$，能否找到 $\boldsymbol{G}\in\mathbf{C}^{n\times m}$，使得 $\boldsymbol{x}=\boldsymbol{Gb}$ 是方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 的极小最小二乘解？如果 $\boldsymbol{G}$ 存在，则 $\boldsymbol{G}$ 有什么特征？

定理 8.6.1 对任意 $\boldsymbol{b}\in\mathbf{C}^m$，$\boldsymbol{x}=\boldsymbol{A}_m^-\boldsymbol{A}\boldsymbol{A}_l^-\boldsymbol{b}$ 是方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 的极小最小二乘解.

证 由定理 8.5.5 知，方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二乘解是

$$\boldsymbol{x} = \boldsymbol{A}_l^- \boldsymbol{b} + (\boldsymbol{E}_n - \boldsymbol{A}_l^- \boldsymbol{A})\boldsymbol{y} \quad , \quad \boldsymbol{y} \in \mathbf{C}^n$$

而且

$$\boldsymbol{Ax} = \boldsymbol{A}\boldsymbol{A}_l^- \boldsymbol{b} + \boldsymbol{A}(\boldsymbol{E}_n - \boldsymbol{A}_l^- \boldsymbol{A})\boldsymbol{y} = \boldsymbol{A}\boldsymbol{A}_l^- \boldsymbol{b}$$

因此对任意 $\boldsymbol{b}\in\mathbf{C}^m$，方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二乘解都是方程 $\boldsymbol{Ax}=\boldsymbol{A}\boldsymbol{A}_l^-\boldsymbol{b}$ 的解. 它的极小范数解为 $\boldsymbol{x}=\boldsymbol{A}_m^-\boldsymbol{A}\boldsymbol{A}_l^-\boldsymbol{b}$，因此 $\boldsymbol{x}=\boldsymbol{A}_m^-\boldsymbol{A}\boldsymbol{A}_l^-\boldsymbol{b}$ 是方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 的极小最小二乘解.

定理 8.6.2 对 $A\in C^{m\times n}$，$G=A_m^-AA_l^-$ 满足 Penrose-Moore 方程

(1) $AGA=A$；

(2) $GAG=G$；

(3) $(AG)^H=AG$；

(4) $(GA)^H=GA$.

证 $AGA=AA_m^-AA_l^-A=AA_l^-A=A$ ；

$GAG=A_m^-AA_l^-AA_m^-AA_l^-=A_m^-AA_m^-AA_l^-=A_m^-AA_l^-=G$；

$(AG)^H=(AA_m^-AA_l^-)^H=(AA_l^-)^H=AA_l^-=AA_m^-AA_l^-=AG$.

$(GA)^H=(A_m^-AA_l^-A)^H=(A_m^-A)^H=A_m^-A=A_m^-AA_l^-A=GA$；

定义 8.6.1 对 $A\in C^{m\times n}$，$G\in C^{n\times m}$，若 G 满足如下 Penrose-Moore 方程

(1) $AGA=A$；

(2) $GAG=G$；

(3) $(AG)^H=AG$；

(4) $(GA)^H=GA$.

则称 G 是 A 的加号逆，或 Penrose-Moore 广义逆，或极小最小二乘广义逆，记为 A^+.

由于 A^+ 满足全部 Penrose-Moore 方程，它都是其他类广义逆的一个，因此它在广义逆中起着重要作用. 下面我们来介绍它的性质以及计算方法.

定理 8.6.3 A 的加号逆唯一存在，且 $A^+=A_m^-AA_l^-$.

证 由定理 8.6.2，$A_m^-AA_l^-$ 是 A 的加号逆.

设 G，F 是 A 的加号逆，则

$$G=GAG=G(AG)^H=G(AFAG)^H=G(AG)^H(AF)^H=GAGAF=GAF$$

$$F=FAF=(FA)^HF=(FAGA)^HF=(GA)^H(FA)^HF=GAFAF=GAF$$

故 $G=F$，即 A 的加号逆唯一. 因此 $A^+=A_m^-AA_l^-$.

定理 8.6.4 设 $A\in C^{m\times n}$，rank $(A)=r>0$，$A=BC$ 是 A 的满秩分解，则

$$A^+=C^H(CC^H)^{-1}(B^HB)^{-1}B^H$$

证 容易验证，$G=C^H(CC^H)^{-1}(B^HB)^{-1}B^H$ 满足 Penrose-Moore 方程，因此 $G=C^H(CC^H)^{-1}(B^HB)^{-1}B^H$ 是 A 的加号逆. 由 A 的加号逆的唯一性知

$$A^+=C^H(CC^H)^{-1}(B^HB)^{-1}B^H$$

下面，我们介绍一些加号逆的性质.

定理 8.6.5 设 A 是 $m\times n$ 矩阵，k 是常数，则

(1) rank (A^+) $=$rank (A)；

(2) AA^+，A^+A 都是幂等阵，且 rank (AA^+) $=$rank (A^+A) $=$rank (A)；

(3) $(A^HA)^+=A^+(A^H)^+=A^+(A^+)^H$；

(4) $A^+=(A^HA)^+A^H=A^H(AA^H)^+$；

(5) $A=AA^H(A^H)^+=(A^H)^+A^HA$；

(6) $(kA)^+=\begin{cases}\dfrac{1}{k}A^+, & k\neq 0\\ 0, & k=0\end{cases}$；

(7) $(A^+)^H=(A^H)^+$；

(8) $(\boldsymbol{A}^{+})^{+}=\boldsymbol{A}$;

(9) $(\boldsymbol{UAV})^{+}=\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}^{+}\boldsymbol{U}^{\mathrm{H}}$，其中 $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}=\boldsymbol{E}_m$，$\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V}=\boldsymbol{E}_n$;

(10) $\boldsymbol{AA}^{+}=(\boldsymbol{AA}^{\mathrm{H}})(\boldsymbol{AA}^{\mathrm{H}})^{+}=(\boldsymbol{AA}^{\mathrm{H}})^{+}(\boldsymbol{AA}^{\mathrm{H}}),\boldsymbol{A}^{+}\boldsymbol{A}=(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}=(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$;

(11) $\boldsymbol{R}(\boldsymbol{A}^{+})=\boldsymbol{R}(\boldsymbol{A}^{\mathrm{H}}),\boldsymbol{N}(\boldsymbol{A}^{+})=\boldsymbol{N}(\boldsymbol{A}^{\mathrm{H}})$.

证 （略）.

定理 8.6.6 rank $(\boldsymbol{A})=tr(\boldsymbol{AA}^{+})=tr(\boldsymbol{A}^{+}\boldsymbol{A})$.

证 因为 $\boldsymbol{AA}^{+}\boldsymbol{AA}^{+}=\boldsymbol{AA}^{+}$，$\boldsymbol{A}^{+}\boldsymbol{AA}^{+}\boldsymbol{A}=\boldsymbol{A}^{+}\boldsymbol{A}$，所以 $\boldsymbol{AA}^{+}$、$\boldsymbol{A}^{+}\boldsymbol{A}$ 是幂等阵，$\boldsymbol{AA}^{+}$、$\boldsymbol{A}^{+}\boldsymbol{A}$ 的特征值只可能是 1 或 0，且 $\boldsymbol{AA}^{+}$、$\boldsymbol{A}^{+}\boldsymbol{A}$ 都可以对角化. 因此

$$\mathrm{rank}(\boldsymbol{A})=tr(\boldsymbol{AA}^{+})=tr(\boldsymbol{A}^{+}\boldsymbol{A}).$$

定理 8.6.7 设 $\boldsymbol{A}\in\mathbf{R}^{m\times n}$，rank $(\boldsymbol{A})=r$，$\boldsymbol{A}=\boldsymbol{PRQ}$ 是 $\boldsymbol{A}$ 的正交三角分解，则

$$\boldsymbol{A}^{+}=\boldsymbol{Q}^{\mathrm{T}}\begin{pmatrix}\boldsymbol{R}_{11}^{-1} & 0\\ 0 & 0\end{pmatrix}\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{R}_{11}^{-1}\boldsymbol{P}_1^{\mathrm{T}}$$

式中：$\boldsymbol{P}$，$\boldsymbol{Q}$ 是正交矩阵；$\boldsymbol{P}_1$，$\boldsymbol{Q}_1$ 分别是 $\boldsymbol{P}$，$\boldsymbol{Q}$ 的前 r 列；$\boldsymbol{R}=\begin{pmatrix}\boldsymbol{R}_{11} & 0\\ 0 & 0\end{pmatrix}$.

证 由定理 8.6.5 (9)，并注意到 $\boldsymbol{R}^{+}=\begin{pmatrix}\boldsymbol{R}_{11}^{-1} & 0\\ 0 & 0\end{pmatrix}$ 即得.

定理 8.6.8 设 $\boldsymbol{A}\in\mathbf{R}^{m\times n}$，rank $(\boldsymbol{A})=r$，$\boldsymbol{A}$ 的奇异值为 σ_1，…，σ_r. 记

$$\boldsymbol{\Lambda}=\mathrm{diag}(\sigma_1^2,\cdots,\sigma_r^2)$$

则 $\boldsymbol{A}^{+}=\boldsymbol{U}\Lambda^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$，其中 $\boldsymbol{U}=(\boldsymbol{p}_1,\cdots,\boldsymbol{p}_r)$，$\boldsymbol{p}_1$，…，$\boldsymbol{p}_r$ 分别是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的对应于 σ_1^2，σ_2^2，…，σ_r^2 的两两正交的单位特征向量.

证 由 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的谱分解有，矩阵 $\boldsymbol{U}$ 使得 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{U\Lambda U}^{\mathrm{T}}$，由定理 8.6.5 (9)，

$$\boldsymbol{A}^{+}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{+}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{U\Lambda}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$$

同理，有如下定理.

定理 8.6.9 设 $\boldsymbol{A}\in\mathbf{R}^{m\times n}$，rank $(\boldsymbol{A})=r$，$\boldsymbol{A}$ 的奇异值为 σ_1，…，σ_r. 记

$$\boldsymbol{\Lambda}=\mathrm{diag}(\sigma_1^2,\cdots,\sigma_r^2)$$

则 $\boldsymbol{A}^{+}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V\Lambda}^{+}\boldsymbol{V}^{\mathrm{T}}$，其中 $\boldsymbol{V}=(\boldsymbol{p}_1,\cdots,\boldsymbol{p}_r)$ 是 $\boldsymbol{AA}^{\mathrm{T}}$ 的对应于 σ_j^2 的两两正交的单位特征向量.

利用 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的谱分解，定理 8.6.9 还可得如下定理.

定理 8.6.10 设 a_1，…，a_s 是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的相异特征值，令

$$P_i(\lambda)=\prod_{\substack{k=1\\k\neq i}}^{s}(\lambda-a_k)$$

有

$$P_i(a_i)=\prod_{\substack{k=1\\k\neq i}}^{s}(a_i-a_k)\quad,\quad \boldsymbol{P}_i(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})=\prod_{\substack{k=1\\k\neq i}}^{s}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}-a_k\boldsymbol{E})$$

则有 $\boldsymbol{A}^{+}=\sum\limits_{i=1}^{s}\mu_i\dfrac{\boldsymbol{P}_i(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})}{P_i(a_i)}\boldsymbol{A}^{\mathrm{T}}$，其中 $\mu_i=\begin{cases}\dfrac{1}{a_i}, & a_i\neq 0\\ 0, & a_i=0\end{cases}$.

证 （略）.

引理 8.6.1　设 $A\in C^{m\times n}$，$U\in C^{n\times p}$，$V\in C^{q\times m}$，rank $(A)=r$，$G=U(VAU)^{-}V$. 则

(1) $G\in A\{1\}$ 的充分必要条件是 rank $(VAU)=r$；

(2) $G\in A\{2\}$ 且 $R(G)=R(U)$ 的充分必要条件是 $\text{rank}(VAU)=\text{rank}(U)$；

(3) $G\in A\{2\}$ 且 $N(G)=N(V)$ 的充分必要条件是 $\text{rank}(VAU)=\text{rank}(V)$；

(4) $G\in A\{1,2\}$ 且 $R(G)=R(U)$，$N(G)=N(V)$ 的充分必要条件是

$$\text{rank}(VAU)=\text{rank}(U)=\text{rank}(V)=r.$$

证　(1) 充分性. 因为

$$r=\text{rank}(VAU)\leqslant\text{rank}(AU)\leqslant\text{rank}(A)=r$$

所以 $\text{rank}(VAU)=\text{rank}(AU)=\text{rank}(A)=r$. 但 $R(AU)\subset R(A)$，因此 $R(AU)=R(A)$，故存在矩阵 P，使得 $A=AUP$. 由定理 8.2.6，有

$$AGA=AU(VAU)^{-}VAUP=AUP=A$$

必要性. 因为 $G\in A\{1\}$，所以

$$A=AGA=AGAGA=AU(VAU)^{-}VAU(VAU)^{-}VA$$

故

$$r=\text{rank}(A)\leqslant\text{rank}(VAU)\leqslant\text{rank}(A)=r$$

即

$$\text{rank}(VAU)=\text{rank}(U)$$

(2) 充分性. 因为 rank (VAU) =rank (U)，由定理 8.2.6，有

$$GAU=U(VAU)^{-}VAU=U$$

因此

$$GAG=GAU(VAU)^{-}V=U(VAU)^{-}V=G$$

$$\text{rank}(U)=\text{rank}(GAU)\leqslant\text{rank}(G)$$

又因为

$$R(G)=R(U(VAU)^{-}V)\subset R(U)$$

故 $G\in A\{2\}$ 且 $R(G)=R(U)$

必要性. 由 $G\in A\{2\}$，有

$$GAG=U(VAU)^{-}VAU(VAU)^{-}V$$

又由 $R(G)=R(U)$，得 $\text{rank}(G)=\text{rank}(U)$，因此

$$\text{rank}(G)\leqslant\text{rank}(VAU)\leqslant\text{rank}(U)=\text{rank}(G)$$

(3) 的证明与 (2) 类似，而 (4) 由 (1)、(2)、(3) 即得.

引理 8.6.2　对任意给定的矩阵 A，$G=A^{+}$ 的充分必要条件是

$$G\in A\{1,2\},R(G)=R(A^{H}),N(G)=N(A^{H})$$

证　易验证 $A^{+}\in A\{1,2\},R(A^{+})=R(A^{H}),N(A^{+})=N(A^{H})$. 反之，由 $R(G)=R(A^{H})$ 知，存在矩阵 $U_1\in C^{m\times m}$，使得 $G=A^{H}U_1$.

由 $N(G)=N(A^{H})$ 有 $(N(G))^{\perp}=(N(A^{H}))^{\perp}$，它等价于 $R(G^{H})=R(A)$，因此存在矩阵 $V_1\in C^{n\times n}$ 使得 $G^{H}=AV^{H}$，即 $G=VA^{H}$.

记 $U_2=(A^{+})^{H}A^{+}\in C^{m\times m}$，$V_2=A^{+}(A^{+})^{H}\in C^{n\times n}$，则有 $A^{+}=A^{H}U_2=V_2A^{H}$.

记 $X=G-A^{+}$，$U=U_1-U_2$，$V=V_1-V_2$，则有

$$AXA=AGA-AA^{+}A=A-A=0,X=A^{H}U=VA^{H}.$$

因此

$$(XA)^{H}(XA)=A^{H}U^{H}AXA=0$$

从而 $XA=0$，因此

$$XX^{H}=X(VA^{H})^{H}=XAV^{H}=0$$

故 $X=0$，即 $G=A^{+}$.

由此可得求 A^{+}的如下 Zlobec 公式.

定理 8.6.11 $A^{+}=A^{H}(A^{H}AA^{H})^{-}A^{H}$.

证 因为

$$\mathrm{rank}(A^{H})=\mathrm{rank}(A^{H}A)=\mathrm{rank}(A^{H}AA^{H}A)\leqslant\mathrm{rank}(A^{H}AA^{H})\leqslant\mathrm{rank}(A^{H})$$

在引理 8.6.1 中取 $U=V=A^{H}$，则 $G=A^{H}(A^{H}AA^{H})^{-}A^{H}$ 满足

$$G\in A\{1,2\},R(G)=R(A^{H}),N(G)=N(A^{H})$$

由引理 8.6.2 知，$G=A^{+}$.

引理 8.6.3 对任意 $n\times n$ 阶 Hermite 矩阵 B，恒有

$$BB^{+}=\lim_{a\to0}(B+aE)^{-1}B=\lim_{a\to0}B(B+aE)^{-1}$$

证 因 B 是 Hermite 矩阵，因此 B 的 n 个特征值 λ_1，λ_2，…，λ_n 都是实数，$B+aE$ 的 n 个特征值 λ_1+a，λ_2+a，…，λ_n+a. 对于充分小的 $|a|>0$，使 $\lambda_i+a\neq0$，从而有 $B+aE$ 非奇异.

因 B 是 Hermite 矩阵，故存在酉矩阵 U，使得

$$B=U\Lambda U^{H}$$

其中 $\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$. 注意到，对任意 $n\times n$ 阶 Hermite 矩阵 B，$BB^{+}=B^{+}B$，因此

$$B=BB^{+}B=BBB^{+}=B^{2}B^{+}$$

从而有

$$\begin{aligned}(B+aE)^{-1}B&=(B+aE)^{-1}B^{2}B^{+}=U(\Lambda+aE)^{-1}\Lambda^{2}U^{H}B^{+}\\&=U\mathrm{diag}[(\lambda_1+a)^{-1}\lambda_1^2,(\lambda_2+a)^{-1}\lambda_2^2,\cdots,(\lambda_n+a)^{-1}\lambda_n^2]U^{H}B^{+}\end{aligned}$$

因此

$$\begin{aligned}&\lim_{a\to0}(B+aE)^{-1}B\\&=U\lim_{a\to0}\mathrm{diag}[(\lambda_1+a)^{-1}\lambda_1^2,(\lambda_2+a)^{-1}\lambda_2^2,\cdots,(\lambda_n+a)^{-1}\lambda_n^2]U^{H}B^{+}\\&=U\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)U^{H}B^{+}\\&=BB^{+}\end{aligned}$$

类似可证 $BB^{+}=\lim\limits_{a\to0}B(B+aE)^{-1}$.

定理 8.6.12 设 A 是 $m\times n$ 矩阵，则有

$$A^{+}=\lim_{a\to0}[(A^{H}A+a^{2}E)^{-1}A^{H}]=\lim_{a\to0}[A^{H}(AA^{H}+a^{2}E)^{-1}]$$

证 由于

$$A^{H}=A^{H}(A^{H})^{+}A^{H}=A^{H}(A^{+})^{H}A^{H}=A^{H}(AA^{+})^{H}=A^{H}AA^{+}$$

又因 $A^{H}A$ 是正定或半正定矩阵，故对任意 $|a|>0$，$A^{H}A+a^{2}E$ 都是可逆矩阵，并且

$$\lim_{a\to0}[(A^{H}A+a^{2}E)^{-1}A^{H}]=\lim_{a\to0}[(A^{H}A+a^{2}E)^{-1}A^{H}A]A^{+}$$

由引理 8.6.3 有

$$\lim_{a\to 0}[(A^HA+a^2E)^{-1}A^HA]A^+=(A^HA)(A^HA)^+A^+$$

由定理 8.6.5（10）有，$(A^HA)(A^HA)^+=A^+A$. 因而有

$$\lim_{a\to 0}[(A^HA+a^2E)^{-1}A^H]=A^+AA^+=A^+$$

类似可证，$A^+=\lim\limits_{a\to 0}[A^H(AA^H+a^2E)^{-1}]$.

由上述定理，可得求 A^+ 的如下方法.

方法一：(1) 求 A 的满秩分解 $A=BC$；

(2) $A^+=C^H(CC^H)^{-1}(B^HB)^{-1}B^H$.

特别地：

1) 若 $A=\begin{pmatrix}A_{11} & 0\\ 0 & 0\end{pmatrix}_{m\times n}$，而 A_{11} 非奇异，则 $A^+=\begin{bmatrix}A_{11}^{-1} & 0\\ 0 & 0\end{bmatrix}_{n\times m}$.

2) 若 $A\in C^{m\times n}$ 行满秩，则 $A^+=A^H(AA^H)^{-1}$.

3) 若 $A\in C^{m\times n}$ 列满秩，则 $A^+=(A^HA)^{-1}A^H$.

【例 8.6.1】 设 $A=\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 1\\ 1 & 2 & 3\end{bmatrix}$，求 A^+.

解 $A=\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 1\\ 1 & 2 & 3\end{bmatrix}\sim\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 1\\ 0 & 0 & 0\end{bmatrix}$，因此 A 的满秩分解为

$$A=\begin{bmatrix}1 & 0\\ 0 & 1\\ 1 & 2\end{bmatrix}\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 1\end{pmatrix}$$

$$\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 1\end{pmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\\ 1 & 1\end{bmatrix}=\begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix},\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 2\end{pmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\\ 1 & 2\end{bmatrix}=\begin{pmatrix}2 & 2\\ 2 & 5\end{pmatrix}$$

$$A^+=\begin{bmatrix}1 & 0\\ 0 & 1\\ 1 & 1\end{bmatrix}\begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix}^{-1}\begin{pmatrix}2 & 2\\ 2 & 5\end{pmatrix}^{-1}\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 2\end{pmatrix}=\frac{1}{6}\begin{bmatrix}4 & -2 & 0\\ -3 & 2 & 1\\ 1 & 0 & 1\end{bmatrix}$$

方法二：(1) 求 A 的正交三角分解 $A=PRQ$，其中 P，Q 是正交矩阵，$R=\begin{pmatrix}R_{11} & 0\\ 0 & 0\end{pmatrix}$；

(2) $A^+=Q^T\begin{bmatrix}R_{11}^{-1} & 0\\ 0 & 0\end{bmatrix}P^T=Q_1^TR_{11}^{-1}P_1^T$，$P_1$，$Q_1$ 分别是 P，Q 的前 rank (A) 列.

【例 8.6.2】 设 $A=\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 1\\ 1 & 2 & 3\end{bmatrix}$，求 A^+.

解 因为 A 的正交三角分解为

$$A=\begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}}\\ 0 & \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}}\end{pmatrix}\begin{pmatrix}\sqrt{12} & 0 & 0\\ \sqrt{18}/2 & \sqrt{6}/2 & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}}\end{pmatrix}$$

rank（A）$=2$，因此

$$A^{+}=\begin{pmatrix}\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}}\\ \frac{2}{\sqrt{6}} & 0\end{pmatrix}\begin{pmatrix}\sqrt{12} & 0\\ \sqrt{18}/2 & \sqrt{6}/2\end{pmatrix}^{-1}\begin{pmatrix}\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\end{pmatrix}$$

$$=\begin{pmatrix}\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}}\\ \frac{2}{\sqrt{6}} & 0\end{pmatrix}\begin{pmatrix}\frac{1}{2\sqrt{3}} & 0\\ -\frac{1}{2} & \frac{\sqrt{6}}{3}\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\end{pmatrix}$$

$$=\frac{1}{6}\begin{pmatrix}4 & -2 & 0\\ -3 & 2 & 1\\ 1 & 0 & 1\end{pmatrix}$$

方法三：(1) 求 A 的奇异值为 $\sigma_1,\cdots,\sigma_r$；

(2) 求 AA^{T} 的对应于 σ_j^2 的两两正交的单位特征向量；

(3) 记

$$\Lambda=\mathrm{diag}(\sigma_1^2,\ \cdots,\ \sigma_r^2)$$

则 $A^{+}=A^{\mathrm{T}}V\Lambda^{+}V^{\mathrm{T}}$，其中 $V=(p_1,\ \cdots,\ p_r)$.

【例 8.6.3】 设 $A=\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 1\\ 1 & -1 & 0\end{pmatrix}$，求 A^{+}.

解 $AA^{\mathrm{T}}=\begin{pmatrix}2 & 1 & 1\\ 1 & 2 & -1\\ 1 & -1 & 2\end{pmatrix}$，$AA^{\mathrm{H}}$ 的非零特征值为 3，3.

$$AA^{\mathrm{H}}-3E=\begin{pmatrix}-1 & 1 & 1\\ 1 & -1 & -1\\ 1 & -1 & -1\end{pmatrix}\sim\begin{pmatrix}1 & -1 & -1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}$$

AA^{T} 的对应于特征值 3 的两两正交的单位特征向量是 $\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ 1\\ 0\end{pmatrix}$，$\frac{1}{\sqrt{6}}\begin{pmatrix}1\\ -1\\ 2\end{pmatrix}$，因此

$$V=\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ 0 & \frac{2}{\sqrt{6}}\end{pmatrix},$$

$$\begin{aligned}A^{+}&=A^{T}V\Lambda^{+}V^{T}\\&=\begin{pmatrix}1&0&1\\0&1&-1\\1&1&0\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ 0 & \frac{2}{\sqrt{6}}\end{pmatrix}\begin{pmatrix}\frac{1}{3}&0\\0&\frac{1}{3}\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}&0\\ \frac{1}{\sqrt{6}}&-\frac{1}{\sqrt{6}}&\frac{2}{\sqrt{6}}\end{pmatrix}\\&=\frac{1}{3}\begin{pmatrix}1&0&1\\0&1&-1\\1&1&0\end{pmatrix}\end{aligned}$$

方法四：(1) 求 $A^{T}A$ 的相异特征值 a_1，…，a_s；

(2) 求 $P_i(a_i)=\prod\limits_{\substack{k=1\\k\neq i}}^{s}(a_i-a_k)$，$P_i(A^{T}A)=\prod\limits_{\substack{k=1\\k\neq i}}^{s}(A^{T}A-a_kE)$；

(3) 求 $A^{+}=\sum\limits_{i=1}^{s}\mu_i\dfrac{P_i(A^{T}A)}{P_i(a_i)}A^{T}$，其中 $\mu_i=\begin{cases}\dfrac{1}{a_i}, & a_i\neq 0\\ 0, & a_i=0\end{cases}$.

【例 8.6.4】 设 $A=\begin{pmatrix}1&0&1\\0&1&1\\1&-1&0\end{pmatrix}$，求 A^{+}.

解 $A^{H}A=\begin{pmatrix}2&-1&1\\-1&2&1\\1&1&2\end{pmatrix}$，$A^{H}A$ 的相异特征值为 $\lambda_1=3$，$\lambda_2=0$，从而 $\mu_1=\frac{1}{3}$，$\mu_2=0$.

因为 $P_1(\lambda)=\lambda$，$P_2(\lambda)=\lambda-3$，故 $P_1(A^{H}A)=A^{H}A$，$P_1(3)=3$（因 $\mu_2=0$，无需求 $P_2(A^{H}A)$，$P_2(\lambda_2)=P_2(0)$）.

$$A^{+}=\left[\frac{1}{9}P_1(A^{H}A)+0\right]A^{H}=\frac{1}{9}A^{H}AA^{H}=\frac{1}{3}\begin{pmatrix}1&0&1\\0&1&-1\\1&1&0\end{pmatrix}$$

方法五：利用 Zlobec 公式 $A^{+}=A^{H}(A^{H}AA^{H})^{-}A^{H}$.

【例 8.6.5】 设 $A^{H}=\begin{pmatrix}1&1&0&2\\0&1&1&1\\2&3&1&5\end{pmatrix}$，求 A^{+}.

解 $A^HAA^H=3\begin{pmatrix}12&18&6&30\\7&11&4&18\\31&47&16&78\end{pmatrix}$

$$\begin{pmatrix}12&18&6&30&1&0&0\\7&11&4&18&0&1&0\\31&47&16&78&0&0&1\\1&0&0&0&&&\\0&1&0&0&&&\\0&0&1&0&&&\\0&0&0&1&&&\end{pmatrix}\sim\begin{pmatrix}1&1&0&2&2/3&-1&0\\7&11&4&18&0&1&0\\0&0&0&0&-2&-1&1\\1&0&0&0&&&\\0&1&0&0&&&\\0&0&1&0&&&\\0&0&0&1&&&\end{pmatrix}$$

$$\sim\begin{pmatrix}1&1&0&2&2/3&-1&0\\0&4&4&4&-14/3&8&0\\0&0&0&0&-2&-1&1\\1&0&0&0&&&\\0&1&0&0&&&\\0&0&1&0&&&\\0&0&0&1&&&\end{pmatrix}\sim\begin{pmatrix}1&0&0&0&2/3&-1&0\\0&1&0&0&-7/6&2&0\\0&0&0&0&-2&-1&1\\1&-1&1&-1&&&\\0&1&-1&-1&&&\\0&0&1&0&&&\\0&0&0&1&&&\end{pmatrix}$$

因此

$$(A^HAA^H)^-=\frac{1}{3}\begin{pmatrix}1&-1&1&-1\\0&1&-1&-1\\0&0&1&0\\0&0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\\0&0&0\end{pmatrix}\begin{pmatrix}2/3&-1&0\\-7/6&2&0\\-2&-1&1\end{pmatrix}$$

$$=\frac{1}{18}\begin{pmatrix}11&-18&0\\-7&12&0\\0&0&0\\0&0&0\end{pmatrix}$$

$$A^+=A^H(A^HAA^H)^-A^H=\frac{1}{18}\begin{pmatrix}4&-2&-6&2\\-7&5&12&2\\1&1&0&2\end{pmatrix}$$

方法六：

$$A^+=\lim_{a\to0}[(A^HA+a^2E)^{-1}A^H]=\lim_{a\to0}[A^H(AA^H+a^2E)^{-1}]$$

【例 8.6.6】 设 $A=\begin{pmatrix}1&2&2\\2&4&4\end{pmatrix}$，求 A^+.

解 $AA^H=\begin{pmatrix}9&18\\18&36\end{pmatrix}$

$$(AA^H+a^2E)^{-1}=\begin{pmatrix}9+a^2&18\\18&36+a^2\end{pmatrix}^{-1}=\frac{1}{45a^2+a^4}\begin{pmatrix}36+a^2&-18\\-18&9+a^2\end{pmatrix}$$

$$A^+=\lim_{a\to0}[A^H(AA^H+a^2E)^{-1}]$$

$$=\lim_{a\to 0}\left[\frac{1}{45+a^2}\begin{pmatrix}1&2\\2&4\\2&4\end{pmatrix}\right]=\frac{1}{45}\begin{pmatrix}1&2\\2&4\\2&4\end{pmatrix}$$

加号逆的 MATLAB 的计算方法如下.

```
X=[]
PINV(X,TOL)
```

或

```
Pinv(X)
```

反馈回来的就是 X 的加号逆.例子如下.

```
X=[1 2 3 4; 5 6 7 8; 2 4 6 8]
Dd=pinv(X)
Dd=
   -0.1100    0.2500   -0.2200
   -0.0450    0.1250   -0.0900
    0.0200   -0.0000    0.0400
    0.0850   -0.1250    0.1700
```

习 题 8

8.1 求矩阵 $A=\begin{pmatrix}0&1&-1&-1&1\\0&-2&2&-2&6\\0&1&-1&-2&3\end{pmatrix}$ 的减号逆 A^- 与 A_r^-.

8.2 求下列矩阵的 A^+

(1)$A=\begin{pmatrix}1&2\\0&0\\2&4\end{pmatrix}$; (2)$A=\begin{pmatrix}i&0\\1&i\\0&1\end{pmatrix}$; (3)$A=\begin{pmatrix}1&-1&0&-1\\0&0&2&2\\-1&1&0&1\end{pmatrix}$.

8.3 求如下线性方程组的极小范数解

$$\begin{cases}x_1+2x_2-x_3+2x_4=1\\2x_1+4x_2+x_3+x_4=5\\-x_1-2x_2-2x_3+x_4=-4\end{cases}$$

8.4 求如下线性方程组的最小二乘解

$$\begin{cases}x_1+2x_2-x_3+2x_4=1\\2x_1+4x_2+x_3+x_4=5\\-x_1-2x_2-2x_3+x_4=-4\end{cases}$$

8.5 求矩阵 $A=\begin{pmatrix}1&1&1&1\\3&2&1&-3\\0&1&2&6\end{pmatrix}$ 的加号逆,并求方程 $Ax=\begin{pmatrix}1\\2\\3\end{pmatrix}$ 的极小最小二乘解.

8.6 设 A 是幂等 Hermite 矩阵,试证明 $A^+=A$.

8.7 证明:$(A\otimes B)^+=A^+\otimes B^+$.

参 考 文 献

1. 徐英凯，李国东，邱启荣. 矩阵理论及其应用. 北京：中国电力出版社，2003
2. 黄有度. 矩阵论及其应用. 合肥：中国科学技术大学出版社，1995
3. 张凯院. 矩阵论典型题解析及自测试题. 西安：西北工业大学出版社，2001
4. 张凯院. 矩阵论同步学习辅导：习题与试题精解. 西安：西北工业大学出版社，2002
5. 徐仲. 理、工科线性代数　常见题型解析及模拟题. 西安：西北工业大学出版社，2002
6. 李尚志编著．线性代数(数学专业用). 北京：高等教育出版社，2006
7. 韩京清，何关钰，许可康. 线性系统理论代数基础[M]. 辽宁：辽宁科学技术出版社，1985
8. 方保镕，周继东，李医民编著. 矩阵论. 北京：清华大学出版社，2004
9. 陈敏，张步涵，段献忠，胡德峰. 基于最小奇异值灵敏度的电压稳定薄弱节点研究. 电网技术第 30 卷第 24 期